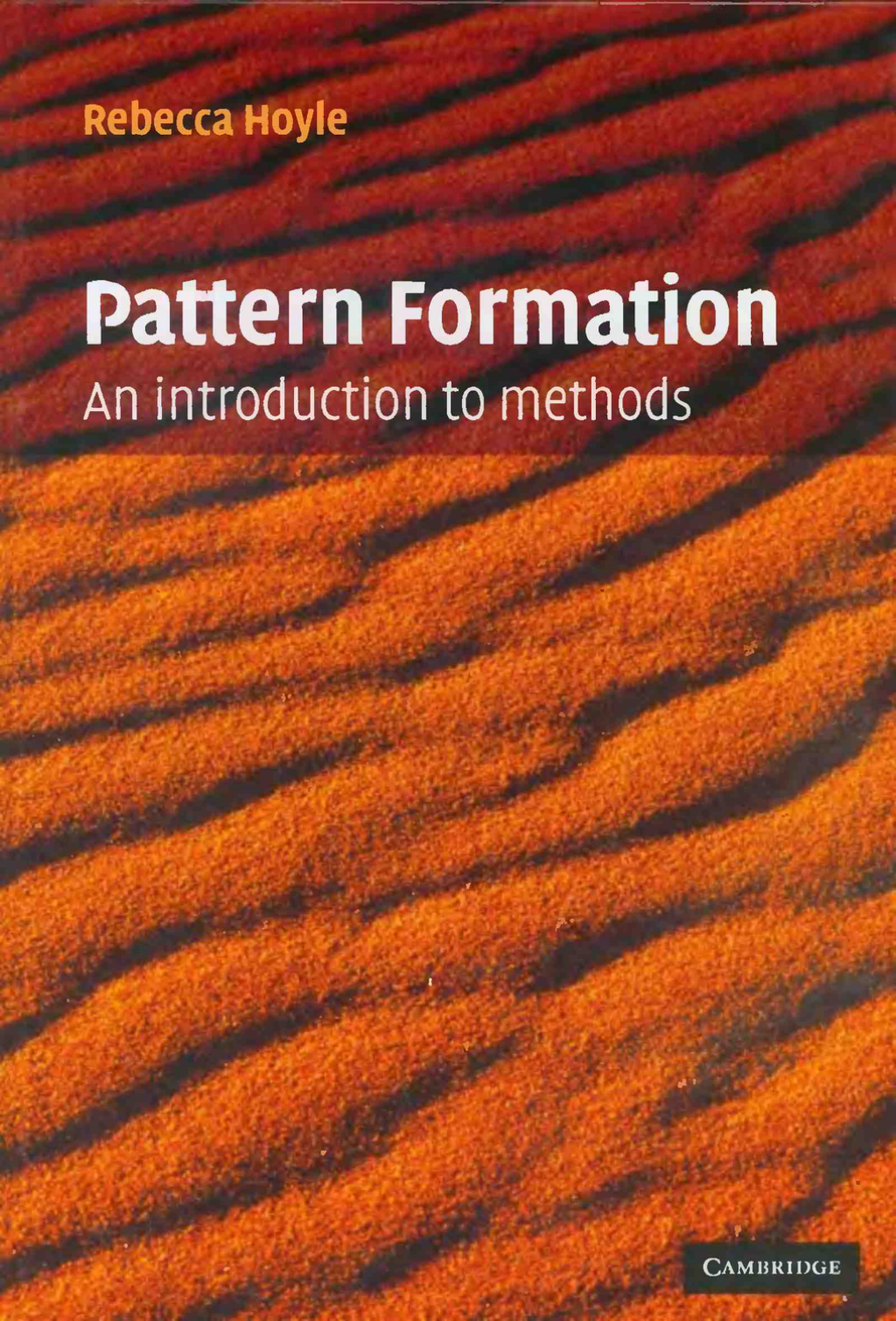
Rebecca Hoyle

Pattern Formation

An introduction to methods



CAMBRIDGE

PATTERN FORMATION

An Introduction to Methods

From the stripes of a zebra and the spots on a leopard's back to the ripples on a sandy beach or desert dune, regular patterns arise in nature everywhere. The appearance and evolution of these phenomena have been a focus of recent research activity across several disciplines.

This book provides an introduction to the range of mathematical theory and methods used to analyse and explain these often intricate and beautiful patterns. Bringing together several different approaches, from group theoretic methods to envelope equations and the theory of patterns in large-aspect-ratio systems, the book provides insight behind the selection of one pattern over another.

Suitable as an upper-undergraduate textbook for mathematics students or as a fascinating, engaging, and fully illustrated resource for readers in physics and biology, Dr Hoyle's book, using a nonpartisan approach, unifies a multiplicity of techniques used by active researchers in this growing field.

REBECCA HOYLE is a Senior Lecturer in Mathematics at the University of Surrey

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York. Melbourne. Madrid, Cape Town, Singapore São Paulo

Cambridge University Press
The Edinburgh Building. Cambridge CB2 2RU. UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge org/9780521817509

First published 2006

Printed in the United Kingdom at the University Press. Cambridge

A catalogue record for this publication is available from the British Library

ISBN-13 978-0-521-81750-9 hardback
ISBN-10 0-521-81750-1 hardback

Contents

Preface

Regular patterns are found in abundance in nature, from the spots on a leopard's back to the ripples on a sandy beach or desert dune. There has a been a flurry of recent research activity seeking to explain their appearance and evolution, and the selection of one pattern over another has turned out to be an inherently nonlinear phenomenon. My aim in writing this book has been to provide an introduction to the range of methods used to analyse natural patterns, at a level suitable for final year undergraduates and beginning graduate students in UK universities.

The book brings together several different approaches used in describing pattern formation, from group theoretic methods to envelope equations and the theory of patterns in large-aspect-ratio systems. The emphasis is on using symmetries to describe universal classes of pattern rather than restricting attention to physical systems with well-known governing equations, though connections with particular systems are also explored. I have taken a wholeheartedly nonpartisan approach, unifying for perhaps the first time in a textbook a multiplicity of methods used by active researchers in the field.

It was David Crighton who originally suggested I should write this book. I had been lecturing a Cambridge Part III course on pattern formation, and David mentioned in passing that it might be a nice idea to turn my lecture notes into a book. Of course I had no idea what I was letting myself in for, but David was always persuasive and inspirational so naturally I said yes. Several years of sweat and toil later I have finally produced the book, though it bears little resemblance to my Part III course, which is probably just as well. I am only sad that David is no longer here to see the result: he inspired and encouraged so many people, particularly those at the beginning of their careers, and he is sorely missed. Like so many others, I owe him a great debt.

Mike Proctor, my Ph.D. supervisor, first introduced me to pattern formation. Later, it was Mike who gave me the opportunity to lecture the course that led to this book. I am very grateful to him for all his support and encouragement over

the years. I particularly enjoy and admire the great enthusiasm with which he approaches any problem, whether it be a tricky calculation or just a matter of finding the right pub.

My heartfelt thanks go to Jon Dawes, Paul Matthews, Mike Proctor, James Robinson and Alastair Rucklidge, who between them read and commented on the manuscript. Their help has made writing this book a much less lonely task, and greatly improved the final draft.

Much of the book was completed while I was on sabbatical in Cambridge, where Mike Proctor and Nigel Weiss were my hosts. I would like to express my thanks to Nigel for his regular reminders to stop working and go home at night.

I would also like to thank the following people who contributed pictures, encouragement or useful comments along the way: Tom Berger, Eberhard Bodenschatz, Steve Cox, Benoit Dionne, Blas Echebarria, Gerhard Ertl, Stephan Fauve, Jay Fineberg, Oliver Harlen (who showed me how to cook hexagons), Dana Mackey, Paul Matthews, Angus McCarter, Ian Melbourne, Tom Mullin, Sarah Pollicott, Hermann Riecke, Alastair Rucklidge, Björn Sandstede, Mary Silber, Annette Taylor, Steve Tobias, Dawn Tse, Laurette Tuckerman, Ed Webb and the editorial team at CUP.

Finally, I am grateful to Nick for cheering me up when it all seemed overwhelming, for his endless patience while his spare room and kitchen table were buried in bits of paper, and for cooking the hexagons with turmeric and then frying them up with potatoes afterwards – delicious!

Guildford, July 2005

1

What are natural patterns?

This book is about patterns: stripes on tigers, whorls in your fingerprints, ripples in sandy deserts, and hexagons you can cook in your own kitchen. More precisely it will be concerned with fairly regular spatial or spatiotemporal patterns that are seen in natural systems – deserts, fingertips, animal coats, stars – and in laboratory or kitchen experiments. These are structures you can pick out by eye as being special in some way, typically periodic in space (Figure 1.1), at least locally. The most common are stripes, squares and hexagons – periodic patterns that tesselate the plane – and rotating spirals or pulsating targets. Quasipatterns with twelvefold rotational symmetry (Figure 1.2) never repeat in any direction, but they look regular at a casual glance, while spiral defect chaos (Figure 1.3) is disordered on a large scale, but locally its constituent moving spirals and patches of stripes are spatially periodic.

Similar patterns are seen in wildly different natural contexts: for example, zebra stripes, desert sand ripples, granular segregation patterns and convection rolls all look stripy, and they even share the same dislocation defects, where two stripes merge into one (Figure 1.4). Rotating spirals appear in a dish of reacting chemicals and in an arrhythmic human heart. Squares crop up in convection and in a layer of vibrated sand. It turns out to be common for a given pattern to show up in several different systems, and for many aspects of its behaviour to be independent of the small details of its environment. This has led to a symmetry-based approach to the description of pattern formation: from this point of view, patterns are universal, and we can find out nearly everything we need to know about them using only their symmetries and those of their surroundings.

This book is intended as an introduction to these symmetry-based techniques and their relationship with more traditional modelling approaches. Before starting on the universal, however, I am going to talk a bit about the specific, describing the archetypal pattern-forming systems: convection, reaction-diffusion and the Faraday wave experiment.

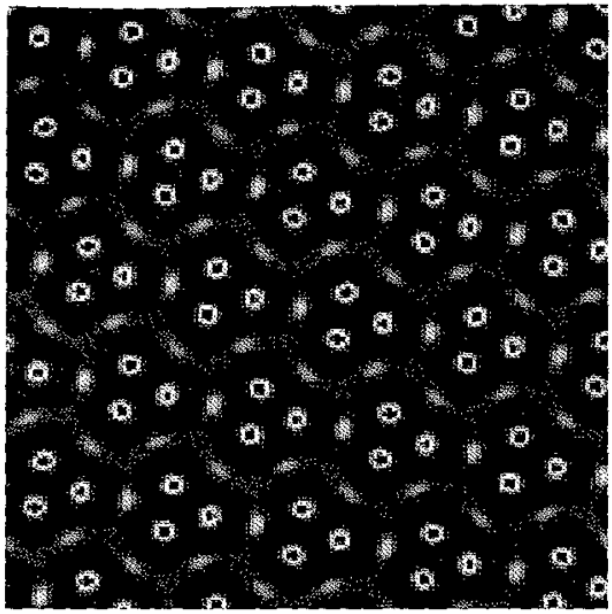


Fig 1 1. A periodic super triangle pattern that tesselates the plane Super triangles can be seen in Faraday wave experiments – see Sections 1.3 and 6.1 and also Silber and Proctor (1998) and Kudrolli, Pier and Gollub (1998) Image courtesy of and ©Mary Silber, Northwestern University, 2003

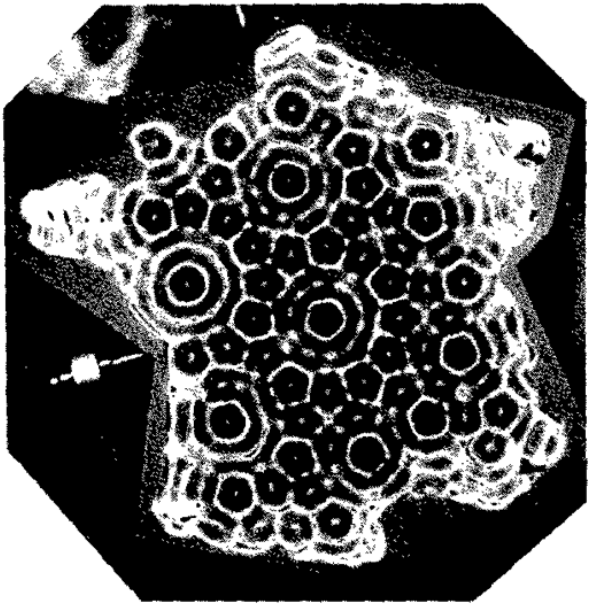


Fig 1 2. Quasipatterns in a Faraday wave experiment. The experimenters chose a container in the shape of France to show that the quasipattern was not caused by boundary effects Reprinted with permission from W S Edwards and S. Fauve, *Physical Review E*, **47**, R788, 1993 ©American Physical Society, 1993

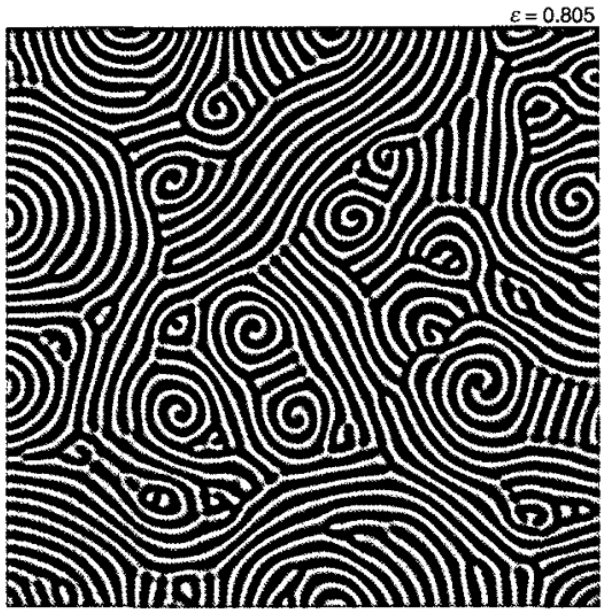


Fig. 1.3 Spiral defect chaos in a Rayleigh–Bénard convection experiment. Image courtesy of and ©Nonlinear Phenomena Group, LASSP, Cornell University, August 2004.

Many of the mathematical techniques and ideas I shall touch upon here are revisited in greater detail in subsequent chapters, so don't worry if you don't follow every step on a first reading. It is enough to get a flavour of the applications to which the theory of pattern formation is relevant. If you are not familiar with simple bifurcation theory it may help to read through the basic ideas in Chapter 2 before attempting to follow the details of the calculations. Simple vector calculus is also needed here, and occasionally in the rest of the book. The descriptions of the phenomena themselves, however, require no particular background knowledge.

Throughout the book I shall use bold italic font for vectors, $\boldsymbol{v}$, but standard italic font for vector-valued functions, $f(t) = v$, and for matrices, scalars and scalar-valued functions.

1.1 Convection

A huge proportion of the early work on pattern formation was motivated by the study of **convection**, which is the overturning of a fluid that is heated from below. Heat at the bottom of a container causes the fluid there to expand, become less dense and more buoyant and so to rise through the colder fluid above. As the

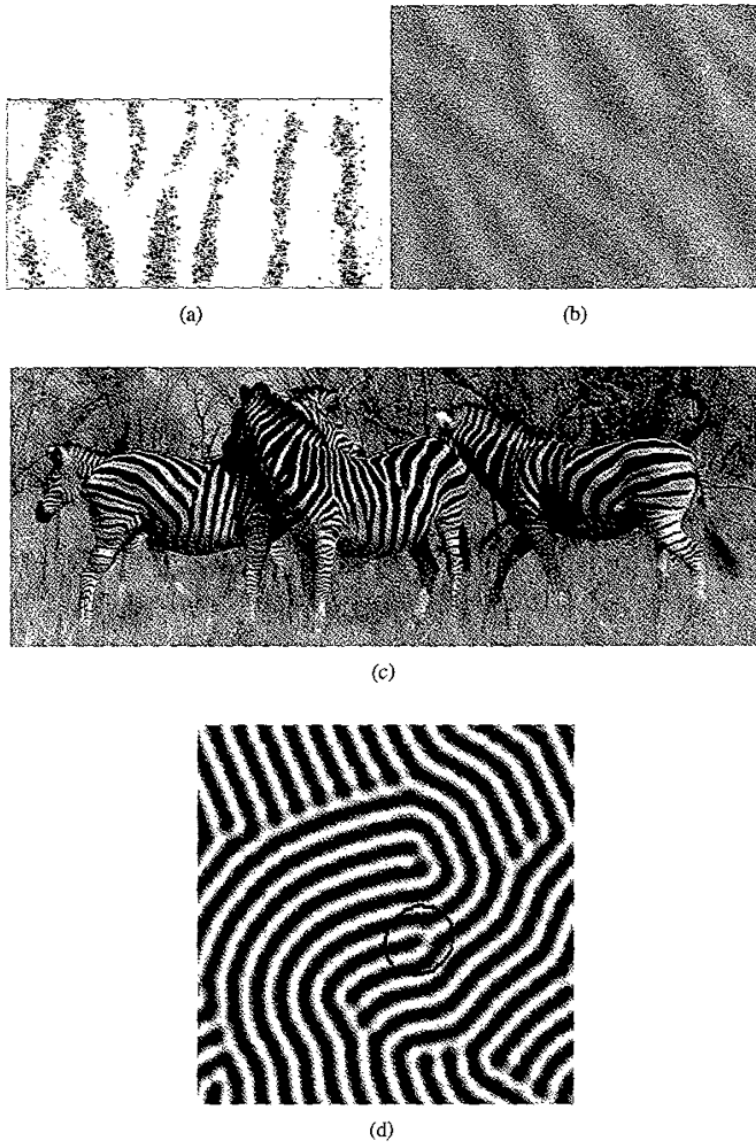


Fig 1 4 Stripe patterns showing dislocations, where two stripes merge into one: (a) segregation in a layer of horizontally shaken sugar and hundreds and thousands (otherwise known as sprinkles or cake decorations); (b) sand ripples in the Sahara desert; (c) on zebras (courtesy of and ©Ed Webb, 2004), and (d) in a numerical simulation of the Swift–Hohenberg convection model Image (a) reprinted with permission from Mullin, T., *Science* **295**, 1851 (2002). ©AAAS (2002)

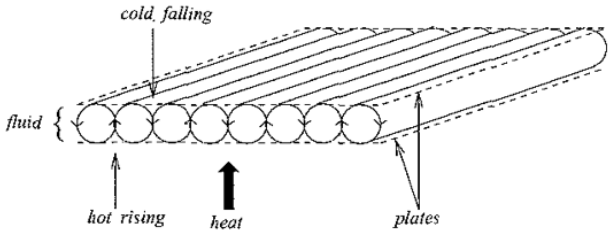


Fig. 1.5 The rise and fall of fluid in convection creates patterns, such as the rolls or stripes shown here. Arrows show the direction of fluid movement. The pattern looks like stripes when observed from the top.

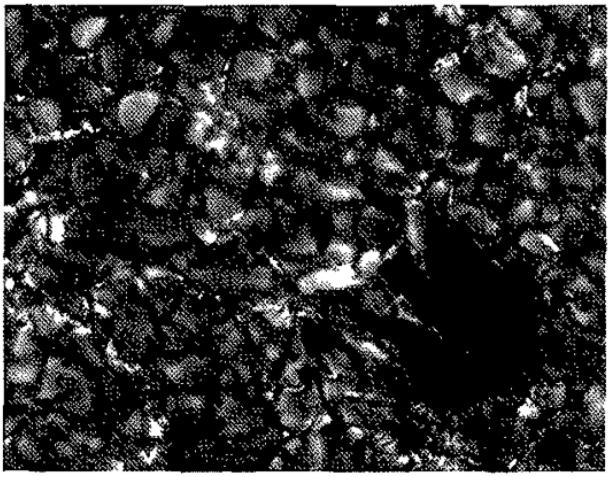


Fig. 1.6. Convection cells in the photosphere of the Sun (solar granulation). The dark region is a sunspot. Image courtesy of and ©Dr. Tom Berger, Lockheed Martin Solar and Astrophysics Lab, Palo Alto, California, 2003.

fluid rises away from the heat source, it cools, becoming denser than the fluid below, and so falls back down to the bottom of the container under the influence of gravity (Figure 1.5). The cycle then repeats, so the fluid is constantly overturning. The rising and falling fluid forms spatial patterns, most commonly **stripes** or **convection rolls** (Figure 1.5), though more complicated patterns such as hexagons and squares are also possible, depending on the details of the physical system and the fluid properties. Convection is often investigated through carefully designed laboratory experiments, but the reason it is so important and has been studied so extensively is that convection occurs naturally in the environment: in the Earth's mantle, convection leads to the movement of tectonic plates or 'continental drift'; in the oceans it drives circulations such as the Gulf Stream that keeps northwestern Europe so much warmer than its northern latitudes would suggest; in the atmosphere, convection creates thunderclouds and in stars, such as the Sun (Figure 1.6),

convection transports energy efficiently from the core where it is produced to the surface where it is released.

In the laboratory, the pattern or **planform** is typically visualised using the **shadowgraph** technique. In this method, a light is shone down onto the convection cell, which must have a transparent top plate and a reflective bottom plate. The warm rising fluid has a lower index of refraction than the cold falling fluid, and so the light is focused towards the cold regions, which appear bright, while the warmer regions remain dark. The pattern can be seen reflected off the bottom plate. Other methods of visualisation are possible, as we shall see in the following kitchen experiment.

1.1.1 How to cook hexagons in your own kitchen

I used to think that apart from stripes, which you can clearly see in fingerprints and on zebras and so on, natural patterns were actually quite exotic – only to be found on the surface of the Sun, and in labs where long hours had been spent in perfecting the experimental set-up. Then I learned how to cook hexagons using only a frying pan, some cooking oil and a sprinkling of pepper.

Cooked hexagons

Warning: This recipe involves hot oil, which is potentially quite dangerous. Only competent adult cooks should attempt to cook hexagons. Do not let any water get into the oil. If the oil starts to smoke, remove the pan from the heat immediately.

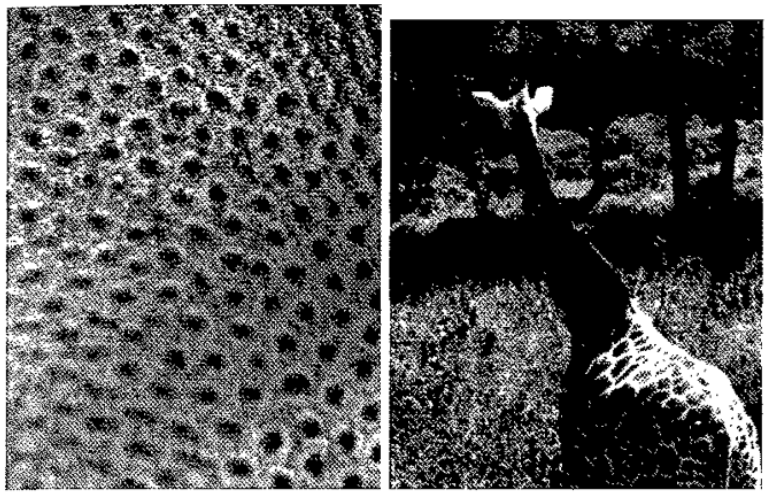
(i) Put a little cooking oil into a flat-bottomed cooking pan. A depth of 0.5–1.00 mm is adequate. You will be able to see the hexagons more easily if the inside of the bottom of the pan is a pale colour. Copper-bottomed pans make the best hexagons because they conduct heat well.

(ii) Mix some very finely ground black pepper or other coloured spice into the oil for visualisation purposes. There should be enough pepper or spice to finely coat the bottom of the pan.

(iii) Put the pan on a flat even heat source – an old-fashioned oil- or coal-fired stove with solid flat plates is best. Gas or electric rings will also work, but the hexagons will be less regular because the heat will be more localised and because the pan is likely to be tilted a bit.

(iv) Heat very gently. Do not let the oil get very hot. A few seconds' heating should be adequate. (Let the hot plate or electric ring heat up first before you put the pan of oil on it.)

(v) Look sideways at the surface of the oil: you should see hexagon-shaped dimples as the oil heats up and starts to convect. You should also see pepper or spice swept along the bottom of the pan into little heaps arranged approximately hexagonally. Once the



(a) (b)

Fig 1.7 Irregular hexagonal patterns in (a) heated cooking oil, and (b) a giraffe's coat markings. Cooked hexagon image courtesy of and ©Nick Safford, 2004

hexagons have formed, the heaps of spice should remain visible if you remove the pan carefully from the heat. In any case, you should not continue to heat the oil for more than a few seconds

(vi) If your hexagons go wrong, take the pan off the heat, cool it down and start again. The hexagons come out best if the oil is cool to start with, and should be seen within a few seconds of heating

Figure 1.7a shows some hexagons cooked using turmeric for visualisation. You can just about see the cell boundaries around each central blob of turmeric. The hexagons are pretty irregular, since this is not a highly controlled experiment. In fact you are likely to see as many pentagons and heptagons as hexagons; giraffe markings also show irregular hexagonal patterns like these (Figure 1.7b) It is also typical to see stripes in the heated oil if the pan is not quite horizontal and the oil is flowing downhill under gravity in places.

1.1.2 Governing equations for Rayleigh–Bénard convection in the Oberbeck–Boussinesq approximation

In 1916, Lord Rayleigh published a paper analysing convection experiments carried out by Henri Bénard and published in 1900. In fact Rayleigh's theory described convection in a fluid that completely fills the gap between the top and bottom plates of a closed cell, whereas Bénard's experiments had used a container that was open at the top so that the fluid had a free surface. These two situations are

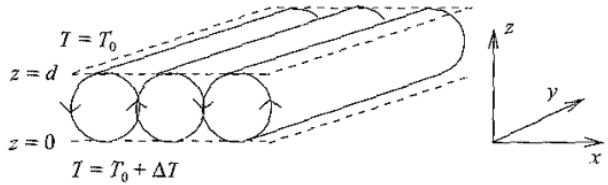


Fig. 1.8 Diagram of the convection system described by equations (1.2) and (1.3). The fluid fills the gap between two horizontal plates at $z = 0$ and d. The top plate is maintained at a temperature $T = T_0$, while the temperature at the bottom is heated to a temperature $T = T_0 + \Delta T$, where $\Delta T > 0$.

actually quite different, because in a filled closed cell buoyancy changes alone are responsible for convection, whereas if the top is open, temperature-induced variations in the surface tension can also drive the motion. Convection between two horizontal plates is known as **Rayleigh–Bénard** or simply **Bénard** convection, while the free surface case is called **Bénard–Marangoni** convection. In his 1958 paper on surface-tension-driven convection Pearson introduced a dimensionless number that measures the relative effects of surface tension and viscous forces; this was later named the Marangoni number after a nineteenth-century Italian scientist, Carlo Marangoni, who noted that fluid flow is coupled to surface tension.

This section will set out the equations used to describe Rayleigh–Bénard convection and show that rolls or stripes are an approximate solution close to onset.

Consider a layer of fluid between two plates at $z = 0$ and d, heated uniformly from below, with the top plate held at a temperature $T = T_0$ and the bottom plate at the higher temperature $T = T_0 + \Delta T$, where ΔT is positive, as shown in Figure 1.8. We assume that the fluid density, ρ, varies linearly with the temperature, T, so that

$$\rho = \rho_0[1 - \alpha(T - T_0)], \tag{1.1}$$

where ρ_0 is the fluid density at $T = T_0$ and α is the (constant) coefficient of thermal expansion, and we further assume that the density variation is only significant in the buoyancy force: this is the **Oberbeck–Boussinesq approximation**. These assumptions are incorporated into the **Navier–Stokes equation** for fluid flow, the heat equation and the continuity equation, to give

$$\rho_0 \left(\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} \right) = -\nabla p - \rho g \hat{\boldsymbol{z}} + \rho_0 \nu \nabla^2 \boldsymbol{u}, \tag{1.2}$$

$$\frac{\partial T}{\partial t} + (\boldsymbol{u} \cdot \nabla) T = \kappa \nabla^2 T, \tag{1.3}$$

$$\nabla \cdot \boldsymbol{u} = 0, \tag{1.4}$$

where $\boldsymbol{u}(x, y, z, t) \in \mathbb{R}^3$ is the three-dimensional fluid velocity, $T(x, y, z, t)$ is the temperature, $p(x, y, z, t)$ is the fluid pressure, g is the (constant) acceleration due to gravity, $\widehat{\boldsymbol{z}}$ is a unit vector in the upward vertical direction, ν is the kinematic viscosity, a measure of the fluid's internal resistance to flow, and κ is the thermal diffusivity that measures the rate of heat conduction through the fluid (see the discussion of diffusion in the following section). Under the Boussinesq approximation, both ν and κ are assumed constant. Details of the derivation of the Navier–Stokes, continuity and heat advection-diffusion equations can be found in any good textbook on fluid dynamics – you might like to try Acheson (1990) if you're interested in finding out more; we will simply accept them as our starting point.

If the heating is not strong enough, the fluid does not convect, but simply conducts heat across the layer. The **conduction solution** is given by

$$\boldsymbol{u} = \boldsymbol{0}, \tag{1.5}$$

$$T = T_c(z) \equiv T_0 + \Delta T \left(1 - \frac{z}{d}\right), \tag{1.6}$$

$$p = p_c(z) \equiv p_0 - \int_0^z \rho(T_c(z)) g \, \mathrm{d}z, \tag{1.7}$$

$$= p_0 - g\rho_0 z \left[1 - \alpha \Delta T \left(1 - \frac{z}{2d}\right)\right], \tag{1.8}$$

where p_0 is the pressure at the bottom of the layer, $z = 0$, and the pressure, $p_c(z)$, is the hydrostatic pressure of fluid in the conducting layer. (The hydrostatic pressure at a height z is the pressure due to the weight of fluid above z.)

When the fluid starts to convect, there will be departures from this conduction solution: to study these, we write $p = p_c(z) + \widehat{p}$ and $T = T_c(z) + \theta$. We also cast the equations into dimensionless form using the substitutions

$$(x, y, z) = d(\widetilde{x}, \widetilde{y}, \widetilde{z}), \tag{1.9}$$

$$t = \frac{d^2}{\kappa}\widetilde{t}, \tag{1.10}$$

$$\boldsymbol{u} = \frac{\kappa}{d}\widetilde{\boldsymbol{u}}, \tag{1.11}$$

$$\theta = \frac{\nu\kappa}{g\alpha d^3}\widetilde{\theta}, \tag{1.12}$$

$$\widehat{p} = \frac{\rho_0 \nu \kappa}{d^2}\widetilde{p} \tag{1.13}$$

The combination of these two sets of substitutions gives

$$\frac{1}{\sigma}\left(\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u}\right) = -\nabla p + \theta\widehat{\boldsymbol{z}} + \nabla^2 \boldsymbol{u}, \tag{1.14}$$

$$\frac{\partial \theta}{\partial t} + (\boldsymbol{u} \cdot \nabla)\theta - Ru_z = \nabla^2 \theta, \tag{1.15}$$

where the tildes ($\tilde{}$) have been dropped immediately to simplify the notation, where u_z is the z-component of $\boldsymbol{u}$ and where $\sigma = \nu/\kappa$ is the **Prandtl number** that measures the relative effects of viscous and thermal diffusion, and

$$R = \frac{\alpha g d^3 \Delta T}{\kappa \nu} \tag{1.16}$$

is the **Rayleigh number** – the nondimensionalised version of the temperature difference between the top and bottom plates.

We now eliminate the pressure by taking the curl of equation (1.14) to get the vorticity equation

$$\frac{1}{\sigma}\left(\frac{\partial \boldsymbol{\omega}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{\omega} - \boldsymbol{\omega} \cdot \nabla \boldsymbol{u}\right) = \nabla\theta \times \hat{\boldsymbol{z}} + \nabla^2 \boldsymbol{\omega}, \tag{1.17}$$

where $\boldsymbol{\omega} = \nabla \times \boldsymbol{u}$ is the fluid vorticity.

To examine the stability of the conduction solution to convection we linearise equations (1.15) and (1.17) around $\boldsymbol{u} = \boldsymbol{\omega} = \boldsymbol{0}, \theta = 0$ giving

$$\frac{1}{\sigma}\frac{\partial \boldsymbol{\omega}}{\partial t} = \nabla\theta \times \hat{\boldsymbol{z}} + \nabla^2 \boldsymbol{\omega}, \tag{1.18}$$

$$\frac{\partial \theta}{\partial t} - R u_z = \nabla^2 \theta \tag{1.19}$$

Now acting on equation (1.18) with $\hat{\boldsymbol{z}} \cdot \nabla \times$ gives

$$\frac{1}{\sigma}\frac{\partial}{\partial t}\nabla^2 u_z = \nabla_h^2 \theta + \nabla^4 u_z, \tag{1.20}$$

where $\nabla_h = (\partial/\partial x, \partial/\partial y, 0)$ is the horizontal gradient operator.

We now need to solve equations (1.19) and (1.20) subject to suitable boundary conditions. The top and bottom plates are held at fixed temperatures, so the temperature perturbation θ must be zero there:

$$\theta = 0, \text{ at } z = 0, 1. \tag{1.21}$$

Mathematically, the simplest velocity boundary conditions to use are the so-called **stress-free** boundary conditions,

$$u_z = \frac{\partial^2 u_z}{\partial z^2} = 0, \text{ at } z = 0, 1, \tag{1.22}$$

that Rayleigh (1916) used in his calculation. We also assume that the convection cell is infinite in horizontal extent so that we do not have to consider any lateral boundary conditions. The solution can now be written as a superposition of Fourier eigenmodes

$$u_z^{(n)}(x, y, z, t) = u_n \sin n\pi z \, e^{i\boldsymbol{k}_h \cdot \boldsymbol{x}_h + st} + c.c., \tag{1.23}$$

$$\theta^{(n)}(x, y, z, t) = \theta_n \sin n\pi z \, e^{i\boldsymbol{k}_h \cdot \boldsymbol{x}_h + st} + c.c., \tag{1.24}$$

where *c.c.* stands for complex conjugate, and where $\boldsymbol{k}_h$ is a horizontal wavevector, $\boldsymbol{x}_h = (x, y, 0)$ is a horizontal position vector, s is the growth rate of the eigenmode, and u_n and θ_n are constants. Substituting one such eigenmode $(u_z^{(n)}, \theta^{(n)})$ into equations (1.19) and (1.20) gives

$$s\theta_n - Ru_n = -(k^2 + n^2\pi^2)\theta_n, \tag{1.25}$$

$$-\frac{1}{\sigma}s(k^2 + n^2\pi^2)u_n = -k^2\theta_n + (k^2 + n^2\pi^2)^2 u_n, \tag{1.26}$$

where $k = |\boldsymbol{k}_h|$. Eliminating the constants θ_n and u_n, gives a dispersion relation

$$s^2(k^2 + n^2\pi^2) + s(1+\sigma)(k^2 + n^2\pi^2)^2 + \sigma(k^2 + n^2\pi^2)^3 - \sigma Rk^2 = 0, \tag{1.27}$$

which shows that the growth rate, s, is zero at

$$R_n(k) = \frac{(k^2 + n^2\pi^2)^3}{k^2} \tag{1.28}$$

In other words, there is a stationary bifurcation at $R = R_n(k)$, (see Chapter 2 for more on bifurcations). The growth rate becomes positive for $R > R_n(k)$ and the nth eigenmode starts to grow. This means that the conduction solution will be unstable to the nth eigenmode $(u_z^{(n)}, \theta^{(n)})$ if $R > R_n(k)$. This happens first for $n = 1$, as $R_n(k)$ is smallest for $n = 1$. The value of k that gives the minimum of

$$R_1(k) = \frac{(k^2 + \pi^2)^3}{k^2}, \tag{1.29}$$

is $k = k_c \equiv \pi/\sqrt{2}$, and this gives the convection instability threshold

$$R_c = R_1(k_c) = \frac{27}{4}\pi^4. \tag{1.30}$$

We expect convection to set in for Rayleigh numbers above this threshold – in other words for a large enough temperature difference between the top and bottom plates.

No-slip velocity boundary conditions, $\boldsymbol{u} = \boldsymbol{0}$, are more realistic than stress-free; they simply say that the fluid must be motionless at the boundaries. The analysis is more complicated in this case, but it is possible to work out a new threshold for convection at $R_c \approx 1708$ (see Manneville (1990) for further details). Lateral boundary conditions can also be accommodated.

Above threshold, any modes that satisfy $R_c \le R_n(k) < R$ can grow. It is straightforward to check that

$$R_n(k) - R_1(k) \ge R_g \equiv 3\pi^4(15 + 2\sqrt{63}), \tag{1.31}$$

for $n \ge 2$. So, close to threshold, where the wavenumber deviation, $\delta k = k - k_c$, and reduced Rayleigh number, $r = (R - R_c)/R_c$, are small and $R < R_c + R_g$,

only the $n = 1$ mode can grow. The growth rate is given by

$$s = \frac{\sigma}{\sigma + 1}\left(\frac{3}{2}\pi^2 r - \delta k^2\right) \tag{1.32}$$

Only wavenumbers in the band $\delta k^2 < 3\pi^2 r/2$ have positive growth rates and the critical wavenumber mode, $k = k_c$ ($\delta k = 0$), grows fastest. Close to the onset of convection we therefore expect to see patterns made up of a superposition of $n = 1$ modes with wavenumbers close to k_c. The wavevectors of the contributing modes can point in any direction, and the linear analysis doesn't tell us anything about which combinations are preferred: the nonlinear terms in the governing equations pick out a small number of modes, as we shall discuss in Chapter 7.

The common convection roll or stripe solution corresponds to the selection of a single pair of wavevectors $\pm\boldsymbol{k}_h$. Wavevectors must occur in equal and opposite pairs, since the velocity and temperature eigenmodes $u_z^{(n)}$ and $\theta^{(n)}$ defined in equations (1.23) and (1.24) must be real. If we choose the x axis to be aligned with $\boldsymbol{k}_h$, then right at onset we have $\boldsymbol{k}_h = (\pi/\sqrt{2}, 0)$ and

$$u_z = u_1 \sin \pi z e^{i\pi x/\sqrt{2}} + c.c., \tag{1.33}$$

$$\theta = \theta_1 \sin \pi z e^{i\pi x/\sqrt{2}} + c.c., \tag{1.34}$$

where one of u_1 and θ_1 is arbitrary at linear order, and the other is then fixed by equation (1.25). Contours of the vertical velocity, u_z, and the temperature perturbation, θ, in the (x, y) plane look like stripes. A regular stripe pattern is shown in Figure 5.1 of Chapter 5. This is what you would see looking down on the convection cell using the shadowgraph technique. Looking from the side, the fluid motion carves out rolls as shown in Figure 1.5.

1.2 Reaction-diffusion systems

Spatial patterns can be seen in systems of reacting and diffusing chemicals. The standard example is the Belousov–Zhabotinsky reaction, where malonic acid is oxidised by bromate ions in the presence of a ferroin catalyst. The reduced state of the catalyst appears red and the oxidised state is blue. Oscillating spiral and target patterns are seen with alternating red and blue arms or rings (Figure 1.9). In fact Belousov (1958) originally used citric acid and a cerium catalyst, where the colour oscillates between yellow and colourless, and Zhabotinsky and his coworkers (see, for example, Zaikin & Zhabotinsky, 1970) extended the work using a variety of other acids and catalysts.

Diffusion is the mechanism by which particles in a fluid are transported from an area of higher concentration to an area of lower concentration through the jostling

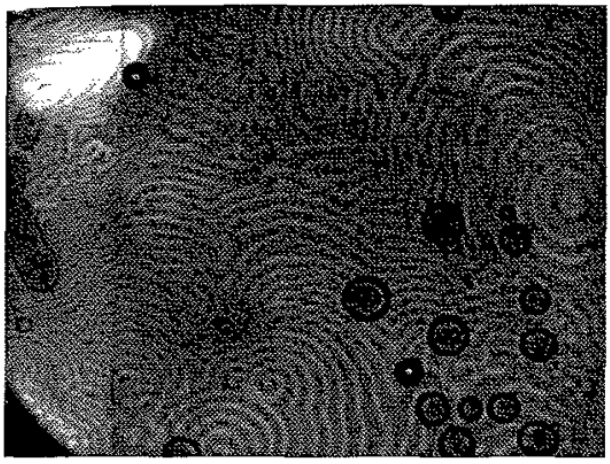


Fig. 1.9 Spirals in the Belousov–Zhabotinsky reaction. Image courtesy of and ©Annette Taylor, University of Leeds, August 2004.

and bumping of the liquid or gas molecules around them, each of which is in constant random thermal motion. Thermal motion takes place on very tiny length-scales and is not apparent to the naked eye, nor are the random movements of any two molecules correlated, so thermal motion does not lead to bulk movement of the fluid. Diffusion can therefore take place in either still or moving fluids. An individual particle is said to be in Brownian motion; diffusion refers to the behaviour of an ensemble of particles. For example, a small amount of smoke released in the corner of a still room will disperse by diffusion. By analogy, heat is said to diffuse through a conductor: if you pick up an ice cube, it is (mostly) diffusion of heat into the ice from your hand that makes your fingers cold.

It might seem paradoxical that diffusion, which tends to smear out high concentrations of a substance and make the distribution of particles more uniform, could possibly lead to pattern formation, where by definition particles of the same type must clump together so that coherent spatial structures can be seen. In the Belousov–Zhabotinsky reaction, and others like it, such as the oxidation of carbon monoxide on the surface of a platinum catalyst (see, for example, Nettesheim *et al*., 1993), the oscillations are caused by the **excitability** of the system, and diffusion simply serves to introduce some local spatial coherence so that neighbouring molecules or parts of the surface oscillate nearly in phase. We will discuss excitable systems in Section 1.2.1. Diffusion can also create patterns more directly: in a famous 1952 paper Turing predicted that two reacting and diffusing chemicals, an **activator** and an **inhibitor**, can produce a pattern if the inhibitor diffuses much faster than the activator. An activator causes growth in the concentration of reactants, whereas an inhibitor causes depletion. It was a very compelling theory, but

it proved very difficult to demonstrate in the laboratory, and it was not until nearly forty years later in 1990 that Castets *et al.* produced the first convincing evidence of an experimental Turing pattern. Turing invoked his reaction-diffusion mechanism to explain **morphogenesis**, the development of shape or form in plants and animals, and it has since been suggested that animal coat markings are created this way, though many biologists strongly dispute it. Classical Turing patterns are steady in time, but oscillatory analogues are possible (see Yang *et al.*, 2002).

Model systems of two reacting and diffusing chemicals, with concentrations $u(\boldsymbol{x}, t)$ and $v(\boldsymbol{x}, t)$, are commonly studied, and are typically written in the form

$$\frac{\partial u}{\partial t} = f(u, v) + D_u \nabla^2 u, \tag{1.35}$$

$$\frac{\partial v}{\partial t} = g(u, v) + D_v \nabla^2 v \tag{1.36}$$

The position vector $\boldsymbol{x}$ can be two- or three-dimensional depending on whether the pattern formation is going on in a thin layer or a large volume. The terms $f(u, v)$ and $g(u, v)$ describe the chemical reactions, while diffusion of u and v is modelled by the terms $D_u \nabla^2 u$ and $D_v \nabla^2 v$, respectively, with D_u and D_v being (positive) diffusion coefficients. We are assuming here that the system is isotropic (invariant under rotations) so that the derivatives in the Laplacian all have the same scaling. We have further assumed homogeneity (invariance under translations) and that D_u and D_v are constants. These equations can describe excitable behaviour or the growth of Turing patterns depending on the form of the reaction terms $f(u, v)$ and $g(u, v)$, and on the values of the diffusion coefficients.

The next two sections will use this basic framework to describe excitability in the FitzHugh–Nagumo equations and the development of Turing patterns respectively.

1.2.1 The FitzHugh–Nagumo model: excitable systems

The FitzHugh–Nagumo equations (FitzHugh, 1961; Nagumo, Arimoto & Yoshizawa, 1962) were originally developed as a model of nerve impulse propagation and are now usually written in the form

$$\frac{\partial u}{\partial t} = u(1 - u)(u + a) - v + I, \tag{1.37}$$

$$\frac{\partial v}{\partial t} = \epsilon(u - bv), \tag{1.38}$$

where I is an external forcing and a, b and $0 < \epsilon \ll 1$ are constants. These are a slight modification of FitzHugh's original equations for a spatially uniform excitation of a nerve axon. Nagumo, Arimoto and Yoshizawa introduced diffusion into

the first equation to describe the movement of the excitation wave along the axon. Variants that include diffusion are now often used as models of general excitable reaction-diffusion systems. We shall consider the modification

$$\frac{\partial u}{\partial t} = \nabla^2 u + f(u, v), \tag{1.39}$$

$$\frac{\partial v}{\partial t} = g(u, v), \tag{1.40}$$

where

$$f(u, v) = \frac{1}{\epsilon} u(1 - u)\left(u - \frac{v + b}{a}\right), \tag{1.41}$$

$$g(u, v) = u - v, \tag{1.42}$$

which Barkley has used extensively to describe spiral waves in excitable media (see, for example, Barkley, 1995, whom we follow in this section). Since ϵ is small, the reaction dynamics of the **excitation variable**, u, are much faster than those of the **recovery variable**, v.

The excitable dynamics comes from the reaction terms: an individual oscillator governed by the same equations without the $\nabla^2 u$ would also behave excitably, so we will ignore the diffusion term for the time being. The origin $u = v = 0$ is a fixed point, $\mathrm{d}u/\mathrm{d}t = \mathrm{d}v/\mathrm{d}t = 0$, of equations (1.39) and (1.40). If $0 < a < 1$ and $b > 0$ it is the only stable fixed point, and has excitable dynamics. To see why, we consider the **nullclines** $f(u, v) = 0$ or $g(u, v) = 0$, plotted in Figure 1.10, which divide the (u, v) phase space into regions according to the signs of $\mathrm{d}u/\mathrm{d}t$ and $\mathrm{d}v/\mathrm{d}t$. We are interested in the region $0 \leq u \leq 1$. A trajectory starting to the left of the line $v = au - b$ decays rapidly towards the origin, since $\mathrm{d}u/\mathrm{d}t$ is large and negative there; v may initially increase, but will decay as soon as the trajectory has crossed the line $v = u$. On the other hand if the initial conditions lie to the right of $v = au - b$ then u will at first grow rapidly away from the origin. However $\mathrm{d}v/\mathrm{d}t$ is also positive, so v will grow, bringing the trajectory back into the region where $\mathrm{d}u/\mathrm{d}t$ is negative and eventually to the origin. This threshold effect is the defining characteristic of an excitable system: small perturbations near to an excitable stable fixed point decay quickly to zero, but disturbances greater than some threshold value lead to large excursions in the dynamics before the stable state is reached. An excitable system is said to be **quiescent** close to the fixed point, **excited** close to the righthand nullcline, here $u = 1$, and **recovering** close to the lefthand nullcline, but far from the fixed point – in this case where $u \approx 0$, but v is large. It is important that recovering states are much further from the threshold, $v = au - b$, than quiescent states, since this means that an excited state must pass through recovery to quiescence before it can be excited again; this is a good model for many

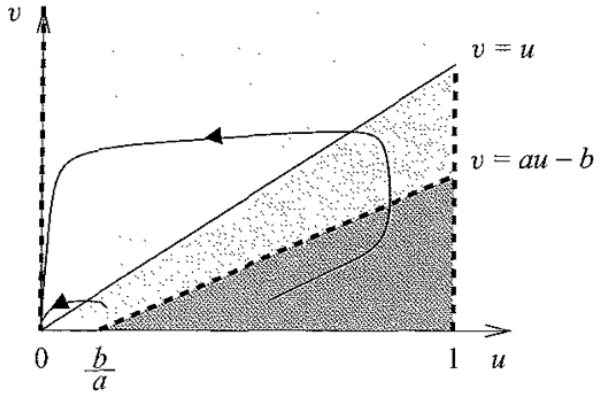


Fig. 1.10. Excitable dynamics in the FitzHugh–Nagumo-type equations (1.39) and (1.40), with $a < 1$ and $b > 0$, in the absence of spatial variation. The u nullclines are given by dashed lines, while the v nullcline is solid. In the light grey region both du/dt and dv/dt are negative; in the medium grey region du/dt is negative and dv/dt is positive, while in the dark grey region both time derivatives are positive. A trajectory starting to the left of the threshold line, $v = au - b$, decays quickly to the stable fixed point at the origin, while a trajectory starting to the right first grows away from the origin, making a large excursion before returning to the fixed point

biological processes that require the slow build-up and rapid discharge of some quantity, such as the action potential in a neuron that FitzHugh and Nagumo, Arimoto and Yoslizawa were originally concerned with

Diffusion couples together the dynamics of neighbouring points in space. In an excitable system this can lead to the propagation of excitation waves. If an excited region, where $u \approx 1$, is next to a quiescent region where u and v are small, the diffusive coupling increases the value of u in the quiescent region (and decreases it in the excited region). Since quiescent points are close in phase space to the excitation threshold, $v = au - b$, this can be sufficient to bring them across the threshold, whereupon the local dynamics take them into excitation. The newly excited region can now excite neighbouring quiescent areas in its turn and a wave of excitation spreads outwards from the initial excited patch. An excited area does not remain excited forever, but goes into the recovery phase, eventually becoming quiescent and ready to be excited again: thus the excitation waves can be periodic like the spiral and target patterns we will discuss in Chapter 10.

The form (1.41) of the reaction term $f(u, v)$ is actually somewhat pathological in that the u nullclines, being three nonparallel straight lines, cross each other. Outside the region $0 \leq u \leq 1$ the dynamics are not those of a simple excitable system. The advantage of the piecewise linear form (1.41) of $f(u, v)$ is that it allows the implementation of a fast numerical scheme for simulating the

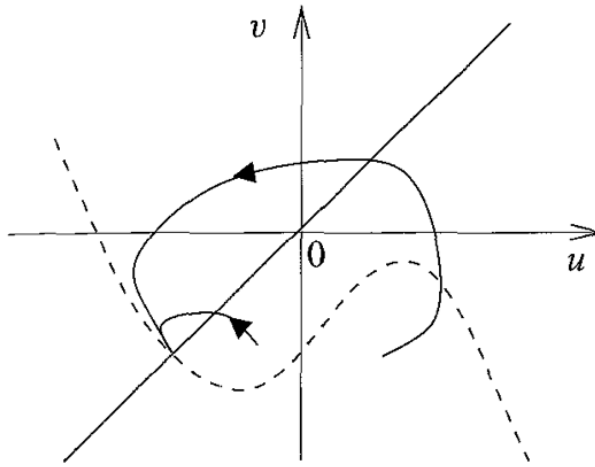


Fig. 1.11 Excitable dynamics in the FitzHugh–Nagumo equations (1.37) and (1.38), for a, b and I all positive and for the case where there is only one stable fixed point at the intersection of the u and v nullclines. The u nullcline is shown as a dashed curve, while the v nullcline is a solid straight line. Two trajectories are shown, one starting in the quiescent region to the left of the middle branch of the u nullcline, and quickly decaying to the fixed point, and one starting to the right of the threshold, and becoming excited before reaching the fixed point.

reaction-diffusion equations (Barkley, 1995). It is also a nice instructive example because it is easy to draw the nullclines, work out where the stable fixed point is, and where the the time derivatives of u and v change sign. The u nullcline of the original FitzHugh–Nagumo equations (1.37) and (1.38) is cubic, and behaves well over the whole range of u, but the fixed points can't be found analytically. In Figure 1.11 the excitable behaviour is sketched in the case where there is only one stable fixed point (at the crossing of the u and v nullclines).

1.2.2 Turing patterns

We now move away from excitability to analyse steady (non-oscillatory) patterns that can arise in reaction-diffusion systems, as predicted by Turing (1952). We start with equations (1.35) and (1.36), and assume that there is a spatially homogeneous solution, $u = u_0$, $v = v_0$, with u_0 and v_0 constants. We now set $u = u_0 + \tilde{u}$, $v = v_0 + \tilde{v}$, where $|\tilde{u}|, |\tilde{v}| \ll 1$ and linearise in $\tilde{u}$ and $\tilde{v}$ to get

$$\frac{\partial \tilde{u}}{\partial t} = a\tilde{u} - b\tilde{v} + D_u \nabla^2 \tilde{u}, \tag{1.43}$$

$$\frac{\partial \tilde{v}}{\partial t} = c\tilde{u} - d\tilde{v} + D_v \nabla^2 \tilde{v}, \tag{1.44}$$

where

$$a = \left.\frac{\partial f}{\partial u}\right|_{(u_0,v_0)}, \quad b = -\left.\frac{\partial f}{\partial v}\right|_{(u_0,v_0)}, \tag{1.45}$$

$$c = \left.\frac{\partial g}{\partial u}\right|_{(u_0,v_0)}, \quad d = -\left.\frac{\partial g}{\partial v}\right|_{(u_0,v_0)}. \tag{1.46}$$

The signs in front of the coefficients are chosen so that for positive a, b, c and d, the chemical corresponding to u is an **activator**, while that corresponding to v is an **inhibitor**. In this context an activator is a chemical that stimulates the growth in concentration of both chemicals, while an inhibitor leads to a decrease in the concentrations. With the signs we have chosen here, positive $\tilde{u}$ (an increase in the concentration, u, of the activator over the steady-state value u_0) gives a positive contribution to the growth of both activator and inhibitor, while positive $\tilde{v}$, an increase in the concentration of inhibitor, leads to a negative contribution to both growth rates.

Turing patterns appear in a diffusion-driven instability where the inhibitor diffuses much faster than the activator, often referred to as **local activation with lateral inhibition**. An initial random perturbation from the steady homogeneous solution is necessary to seed the instability. Any locally high concentration of activator in the perturbation causes both activator and inhibitor concentrations to increase. The inhibitor diffuses away from the area more quickly than the activator, and so the relative concentration of inhibitor becomes high in a region outside the original patch. Neither the activator nor the inhibitor are being produced in this border region, so the activator becomes depleted there. The inhibitor is constantly replenished by diffusion from the central activator-rich patch, but beyond the border region levels start to drop off because there is no local supply. This allows activator to build up again if there is another localised patch of activator in the initial random perturbation, so the pattern can repeat periodically in space (Figure 1.12).

The conditions for a Turing instability to occur can be derived from equations (1.43) and (1.44). We want the instability to be diffusion-driven, so the system should be stable in the absence of diffusion. If we set $\tilde{u} = \hat{u}e^{\sigma t}$ and $\tilde{v} = \hat{v}e^{\sigma t}$, where $\hat{u}$ and $\hat{v}$ are constants, the growth rate eigenvalues are given by

$$\sigma = \frac{1}{2}(a-d) \pm \frac{1}{2}\sqrt{(a+d)^2 - 4bc}. \tag{1.47}$$

The solution $u = u_0$, $v = v_0$, is stable if both eigenvalues are negative, which requires $a < d$ and $ad < bc$.

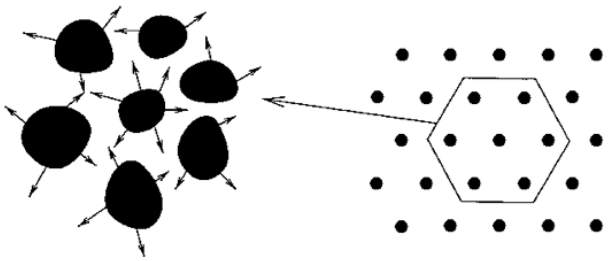


Fig. 1.12. Inhibitor diffuses out of patches of high activator concentration where it is produced, and creates an inhibitory border, resulting in a Turing pattern Regions of higher activator concentration are shown in black while areas of higher inhibitor concentration are coloured white The arrows indicate the diffusion of inhibitor. The area pictured can be considered as part of an array of patches placed approximately at the vertices of an infinite hexagonal lattice in this example.

Any spatially varying solution $(\widetilde{u}(\boldsymbol{x}, t), \widetilde{v}(\boldsymbol{x}, t))$ can be expressed as a Fourier series in space, so we need to know which Fourier modes

$$\widetilde{u} = \widehat{u} e^{i\boldsymbol{k}\cdot\boldsymbol{x}+\sigma t} + c.c \tag{1.48}$$

$$\widetilde{v} = \widehat{v} e^{i\boldsymbol{k}\cdot\boldsymbol{x}+\sigma t} + c.c. \tag{1.49}$$

will grow, where $\boldsymbol{k}$ is a constant wavevector and $\widehat{u}$ and $\widehat{v}$ are complex constants. We shall assume for the moment that the system is unbounded in space so that we don't have to worry about spatial boundary conditions. The necessary modifications to take account of them can be made quite easily later on if required.

Substituting (1.48) and (1.49) into the linearised equations (1.43) and (1.44) leads to the dispersion relation for σ:

$$\sigma^2 + \sigma(D_u k^2 + D_v k^2 - a + d) + (D_u k^2 - a)(D_v k^2 + d) + bc = 0, \tag{1.50}$$

where $k = |\boldsymbol{k}|$. For instability, at least one of the roots of this equation must have positive real part. The sum of the roots is

$$-(D_u k^2 + D_v k^2 - a + d), \tag{1.51}$$

which is negative since $a < d$ and both diffusion coefficients are positive. The only possibility for an instability is therefore to have one negative and one positive real root, which will be the case as long as the product of the roots is negative:

$$h(k^2) \equiv (D_u k^2 - a)(D_v k^2 + d) + bc < 0 \tag{1.52}$$

If the minimum, $h_{\min}$, of the function $h(k^2)$ is negative then a range of modes $k_1^2 < k^2 < k_2^2$ will grow (Figure 1.13). The values of the coefficients in the inequality (1.52) are fixed by the chemistry, and as a result so are $h_{\min}$, k_1 and k_2

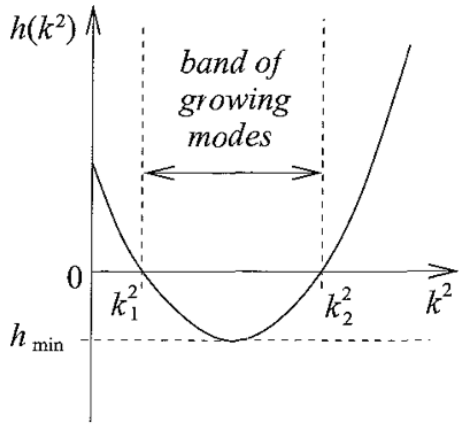


Fig. 1.13. Modes with wavenumbers lying between the zeros, k_1^2 and k_2^2, of $h(k^2)$ grow in the Turing instability.

It is easy to check that the minimum of $h(k^2)$ is

$$h_{\min} = -\frac{1}{4D_uD_v}(dD_u - aD_v)^2 - ad + bc \tag{1.53}$$

and occurs at

$$k^2 = \frac{1}{2}\left(\frac{a}{D_u} - \frac{d}{D_v}\right). \tag{1.54}$$

So the Turing instability occurs as long as $a/D_u > d/D_v$ (so that k^2 is positive) and $h_{\min} < 0$. The lengthscales $l_u = \sqrt{D_u/a}$ and $l_v = \sqrt{D_v/d}$ give a measure of the distance over which u and v, respectively, will decay to low values from peak concentrations. The requirement that k^2 be positive can be rewritten as $l_v > l_u$, which says that the inhibitor must penetrate further than the activator. This is the local activation with lateral inhibition that was discussed earlier.

1.3 Faraday waves

In 1831, Faraday published observations of 'crispations' in vertically vibrated layers of fluid, in other words standing-wave deformations of the fluid surface. These parametrically excited surface waves form a variety of striking patterns: Faraday himself saw square or rectangular grids of wavecrests, or circular patterns, depending on the strength of the vibration. Subsequent experiments have revealed more exotic patterns, such as the one with twelvefold rotational symmetry in Figure 1.2. Faraday used an array of different household fluids – milk, egg white, alcohol, ink and turpentine – to produce his patterns. He also found

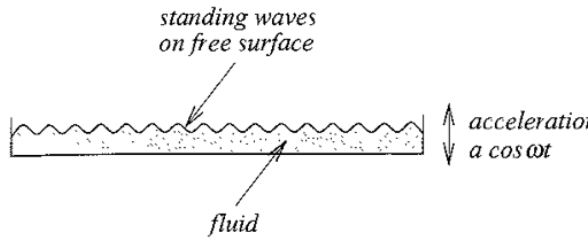


Fig 1.14 A container of fluid is vibrated up and down sinusoidally. Surface waves form in the Faraday instability

crispations of 'extreme beauty' in mercury Recently patterns including stripes, squares and hexagons have been seen in vibrated layers of granular material (see, for example, Melo, Umbanhowar & Swinney, 1995), though in this case the instability mechanism is different from Faraday's Faraday observed his patterns by eye using reflections from the fluid surface or, particularly in the case of milk, by shining a light through the fluid from below He also sprinkled sand over the container and saw that it was arranged into patterns under the water. In modern experiments the patterns are typically imaged by reflection from a light placed near a camera. Flat regions of the surface reflect light directly back to the camera, and so they appear bright, while inclined surfaces appear darker. Usually the images are averaged over a whole cycle of oscillation.

In a typical experiment the fluid layer is vibrated vertically (Figure 1 14) with acceleration $a\cos\omega t$, for constant a and ω, so that the effective gravitational acceleration felt by the layer is

$$g_{\text{eff}}(t) = -g + a\cos\omega t, \tag{1.55}$$

where g is the acceleration due to gravity on a still layer. At low forcing amplitude, a, the surface of the layer remains flat, but standing waves or crispations form once the amplitude increases beyond a threshold value, a_c, at which the forced acceleration is strong enough to overcome viscous dissipation, the loss of energy as heat through fluid friction For deep enough layers the response is subharmonic, the surface waves oscillating at half the driving frequency, as observed by Faraday who wrote that 'each heap . . recurs or is re-formed in two complete vibrations of the sustaining surface'

The driving does not have to be sinusoidal: the pattern shown in Figure 1.2 was produced using two-frequency forcing of the form

$$g_{\text{eff}} = -g + a[\cos\theta\cos 4\omega t + \sin\theta\cos(5\omega t + \phi)] \tag{1 56}$$

for constant a, θ, ω and ϕ. This type of forcing can produce a variety of interesting planforms including examples of so-called superlattices and quasipatterns, which

will be described in detail in Chapter 6. The effective gravity, g_{eff}, is periodic with period $2\pi/\omega$; for appropriate choices of the parameters this forcing can lead to patterns that are harmonic, oscillating with this same period.

The calculation of the instability threshold is complicated, even for the simple forcing (1.55), (see, for example, Kumar, 1996), so I will not present it here. However the starting points are the Navier–Stokes equation, with time-varying effective gravity, and the continuity equation,

$$\rho\left(\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u}\right) = -\nabla p - \rho(g - a\cos\omega t)\hat{\boldsymbol{z}} + \rho\nu\nabla^2\boldsymbol{u}, \tag{1.57}$$

$$\nabla\cdot\boldsymbol{u} = 0, \tag{1.58}$$

together with suitable boundary conditions that prescribe, among other things, how the fluid responds to surface tension at the free surface.

1.4 Outline of the rest of the book

Convection, reaction-diffusion and Faraday waves comprise the three most commonly studied pattern-forming systems, and much of the theory of pattern formation has been developed in an attempt to explain experimental results in one or other of them. Of course, there are others: spatial patterns can be seen in flame fronts, lasers and solidifying metal alloys, for example. The theory set out in the rest of this book will start from the symmetries and observable features of the patterns themselves, rather than the specifics of any one experiment.

Chapters 2 and 3 set out some introductory material on bifurcation theory and group theory, respectively, that will be needed for the analysis of patterns. If you have a strong background in either or both of these areas you can simply skip the relevant chapter(s) and refer back to them if you need to. The bulk of the book from Chapter 4 onwards describes theoretical approaches to understanding pattern formation. Roughly speaking we start with the most regular patterns in Chapters 4 and 5 and work towards the most irregular in Chapters 10 and 11. There is a corresponding transition from the theory of bifurcations with symmetry at the beginning of the book to asymptotic methods at the end.

At the end of each chapter from Chapter 2 onwards there is a set of exercises. A set of partial solutions to the exercises can be found on the web page www.cambridge.org/9780521817509. I will also post any errata at the same address from time to time.

2

A bit of bifurcation theory

The patterns we are going to study in this book arise at bifurcations, so to understand patterns we first need to know something about bifurcations. This chapter is a very brief review of local bifurcation theory, together with a few other useful results from the theory of nonlinear systems. I explain the main ideas that are needed to understand symmetric bifurcations in later chapters, presenting the material informally, without proofs. For readers who are new to all this and would like to read about the material presented here in more depth I recommend the books by Glendinning (1994) and Guckenheimer and Holmes (1983). This chapter follows Glendinning quite closely in places, and his book would be an excellent place to look for further details on any of the topics outlined here.

2.1 Flows, stationary points and periodic orbits

Just as at a fork in the road you suddenly have two paths ahead of you instead of one, so at a bifurcation in a nonlinear system there are sudden changes in flows, stationary points and periodic orbits. Before we go any further we need to define all these things.

Let us start with the ordinary differential equation

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = f(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^n \tag{2.1}$$

where t is time and $\boldsymbol{x}$ is a position vector in $\mathbb{R}^n$, the **phase space** of the differential equation. Throughout this book we shall assume that f is sufficiently smooth that the techniques we use, such as Taylor series expansions, are valid. The solutions of equation (2.1) define a **flow**, $\phi(\boldsymbol{x}, t)$, such that $\boldsymbol{x}(t) = \phi(\boldsymbol{x}_0, t)$, where $\boldsymbol{x}_0$ is the initial value of the solution at $t = 0$ and so labels all the different possible solutions according to their starting position. In other words $\phi(\boldsymbol{x}_0, 0) = \boldsymbol{x}_0$. The

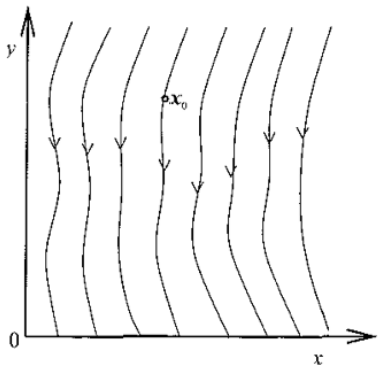


Fig 2 1 An example of a flow in the (x, y) plane. The starting point (at $t = 0$) of one of the trajectories in the flow is labelled x_0 Each trajectory has a different starting point, and so a different value of x_0.

flow is defined for all t, positive and negative, such that the solution through x_0 exists.

In the case of a two-dimensional vector $x = (x, y) \in \mathbb{R}^2$ we can make a two-dimensional picture, or **phase portrait**, of a flow like the one in Figure 2.1. Each line on the diagram is a **trajectory**, a solution of the differential equation labelled by a particular starting point x_0 and parameterised by time. The arrows represent the direction of increasing time. The $t < 0$ sections of each trajectory are drawn by running time backwards The figure makes clear why we say that the solutions define a flow: they look like a snapshot of water flowing over the page, or lines of marching ants

A **stationary point** (or **fixed point** or **equilibrium solution**) of equation (2 1) is a point x such that

$$\frac{dx}{dt} = f(x) = \mathbf{0}. \tag{2 2}$$

The flow in Figure 2.2 has a fixed point at $x = x_1$.

Example 2.1 *The one-dimensional differential equation*

$$\frac{dx}{dt} = x - x^3, \tag{2.3}$$

has fixed points at $x = 0, \pm 1$.

A differential equation may also have periodic orbits A point, x, is **periodic** under equation (2.1) with period T if and only if $\phi(x, t + T) = \phi(x, t)$ for all t and $\phi(x, t + s) \neq \phi(x, t)$ for all $0 < s < T$. This says that the trajectory starting at x at time t first returns to the point x after an additional time T The closed curve $C = \{y | y = \phi(x, t), 0 \leq t < T\}$ is a **periodic orbit**, and consists of the trajectory

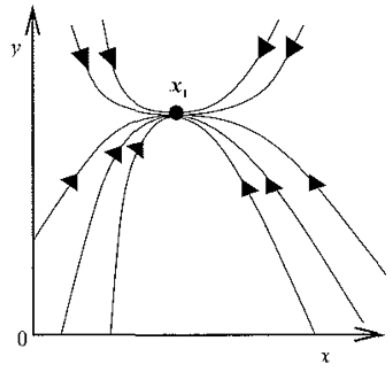


Fig. 2.2 An example of a flow with a fixed point at $x = x_1$.

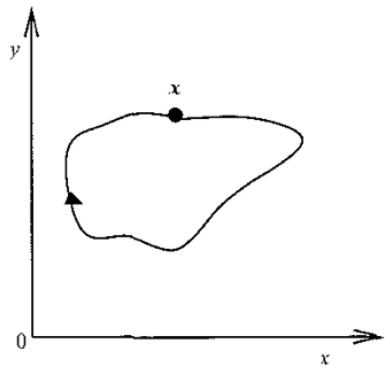


Fig. 2.3 This flow has a periodic orbit passing through the periodic point x. Only the trajectory making up the periodic orbit is shown

joining the periodic point x back to itself in phase space. A flow containing a periodic orbit is shown in Figure 2.3.

Example 2.2 *The equations*

$$\frac{dx}{dt} = x(1 - x^2 - y^2) - y, \tag{2.4}$$

$$\frac{dy}{dt} = y(1 - x^2 - y^2) + x, \tag{2.5}$$

have a periodic orbit of the form $x = \cos t$, $y = \sin t$. *Every point* $x = (x, y)$ *on the orbit is periodic with period* $T = 2\pi$

A stationary point is categorised according to the phase portrait, the picture of the flow in phase space, immediately around it. This is partly determined by the

Jacobian matrix, $Df(\boldsymbol{x})$, defined by

$$Df_{ij} = \frac{\partial f_i}{\partial x_j}, \tag{2.6}$$

evaluated at the stationary point, and in particular by its eigenvalues and eigenvectors.

Eigendirections corresponding to eigenvalues with negative real part are directions in which the stationary point is stable, and those corresponding to eigenvalues with positive real part are unstable directions. To see this we will investigate the stability of the stationary point $\boldsymbol{x} = \boldsymbol{x}_0$ by perturbing it slightly so that $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{\eta}(t)$, with $0 < |\boldsymbol{\eta}| \ll 1$ Substituting for $\boldsymbol{x}$ in equation (2.1) gives

$$\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}t} = \boldsymbol{f}(\boldsymbol{x}_0 + \boldsymbol{\eta}). \tag{2.7}$$

Now Taylor-expanding $\boldsymbol{f}(\boldsymbol{x}_0 + \boldsymbol{\eta})$ in powers of $\boldsymbol{\eta}$ we get

$$\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}t} = \boldsymbol{f}(\boldsymbol{x}_0) + D\boldsymbol{f}\big|_{\boldsymbol{x}=\boldsymbol{x}_0}\boldsymbol{\eta} + O(|\boldsymbol{\eta}|^2) \tag{2.8}$$

Since $\boldsymbol{x}_0$ is a stationary point of equation (2.1) we know that $\boldsymbol{f}(\boldsymbol{x}_0) = \boldsymbol{0}$ If $\boldsymbol{\eta}$ is an eigenvector of $D\boldsymbol{f}\big|_{\boldsymbol{x}=\boldsymbol{x}_0}$ with eigenvalue λ, then $D\boldsymbol{f}\big|_{\boldsymbol{x}=\boldsymbol{x}_0}\boldsymbol{\eta} = \lambda\boldsymbol{\eta}$, so the evolution equation for $\boldsymbol{\eta}$ becomes

$$\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}t} = \lambda\boldsymbol{\eta} + O(|\boldsymbol{\eta}|^2). \tag{2.9}$$

Solving this equation we find that to leading order $\boldsymbol{\eta} = \boldsymbol{\eta}_0 e^{\lambda t}$, where $\boldsymbol{\eta}_0$ is the value of $\boldsymbol{\eta}$ at $t = 0$ Now, if the real part of λ is negative, we find that $|\boldsymbol{\eta}| \to 0$ as $t \to +\infty$ and so the perturbation dies away: in other words the stationary point is stable in this direction. Conversely if the real part of λ is positive, we find that the size of the perturbation, $|\boldsymbol{\eta}|$, grows with time and the stationary point is unstable in the direction of $\boldsymbol{\eta}$

A stationary point is only linearly stable if all its eigenvalues have negative real part: if any eigenvalue has zero real part and none has positive real part the point is linearly neutrally stable, and if any eigenvalue has positive real part the point is linearly unstable

When we are considering motion in the plane ($\boldsymbol{x} \in \mathbb{R}^2$) there are two eigenvalues λ_1 and λ_2 and the different types of stationary point are given names as follows:

(i) $\lambda_1 < \lambda_2 < 0$: a stable node;
(ii) $\lambda_1 = \lambda_2 < 0$ and all vectors are eigenvectors: a stable star;
(iii) $\lambda_1 = \lambda_2 < 0$ and there is only one eigenvector: a stable improper node;
(iv) $\lambda_1 > \lambda_2 > 0$: an unstable node;

(v) $\lambda_1 = \lambda_2 > 0$ and all vectors are eigenvectors: an unstable star;
(vi) $\lambda_1 = \lambda_2 > 0$ and there is only one eigenvector: an unstable improper node;
(vii) $\lambda_1 < 0 < \lambda_2$: a saddle point;
(viii) $\lambda_1 = +i\omega, \lambda_2 = -i\omega$: a centre;
(ix) $\lambda_1 = \rho + i\omega, \lambda_2 = \rho - i\omega, \rho < 0$: a stable focus;
(x) $\lambda_1 = \rho + i\omega, \lambda_2 = \rho - i\omega, \rho > 0$: an unstable focus;
(xi) degenerate cases where an eigenvalue is exactly zero

All but the degenerate cases are shown in Figure 2.4

Example 2.3 *Consider the equations*

$$\frac{dx}{dt} = 3x + x^2, \tag{2.10}$$

$$\frac{dy}{dt} = y + xy \tag{2.11}$$

There is a stationary point at $x = y = 0$. The Jacobian is given by

$$Df = \begin{pmatrix} 3 + 2x & 0 \\ y & 1 + x \end{pmatrix}. \tag{2.12}$$

Evaluating the matrix at the stationary point $x = y = 0$ gives

$$Df\big|_{x=y=0} = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.13}$$

The eigenvalues are $\lambda_1 = 3$, with corresponding eigenvector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\lambda_2 = 1$, with eigenvector $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Perturbing around $x = y = 0$ by setting $x = \eta_x$ and $y = \eta_y$, where $|\eta_x| \ll 1$ and $|\eta_y| \ll 1$ and linearising, we find the following equations for the growth of the perturbations:

$$\frac{d\eta_x}{dt} = 3\eta_x, \tag{2.14}$$

$$\frac{d\eta_y}{dt} = \eta_y \tag{2.15}$$

Integrating these we find $\eta_x = \widehat{\eta}_x e^{3t}$ and $\eta_y = \widehat{\eta}_y e^t$, where $\widehat{\eta}_x$ and $\widehat{\eta}_y$ are the values of the perturbations at time $t = 0$. The perturbations grow as time increases, so the stationary point at $x = y = 0$ is unstable. According to the classification above, the fixed point is in fact an unstable node. The phase portrait in a neighbourhood of the origin is shown in Figure 2.5, since the full nonlinear equations were used to produce it, the picture looks slightly different from the linearised version presented in Figure 2.4.

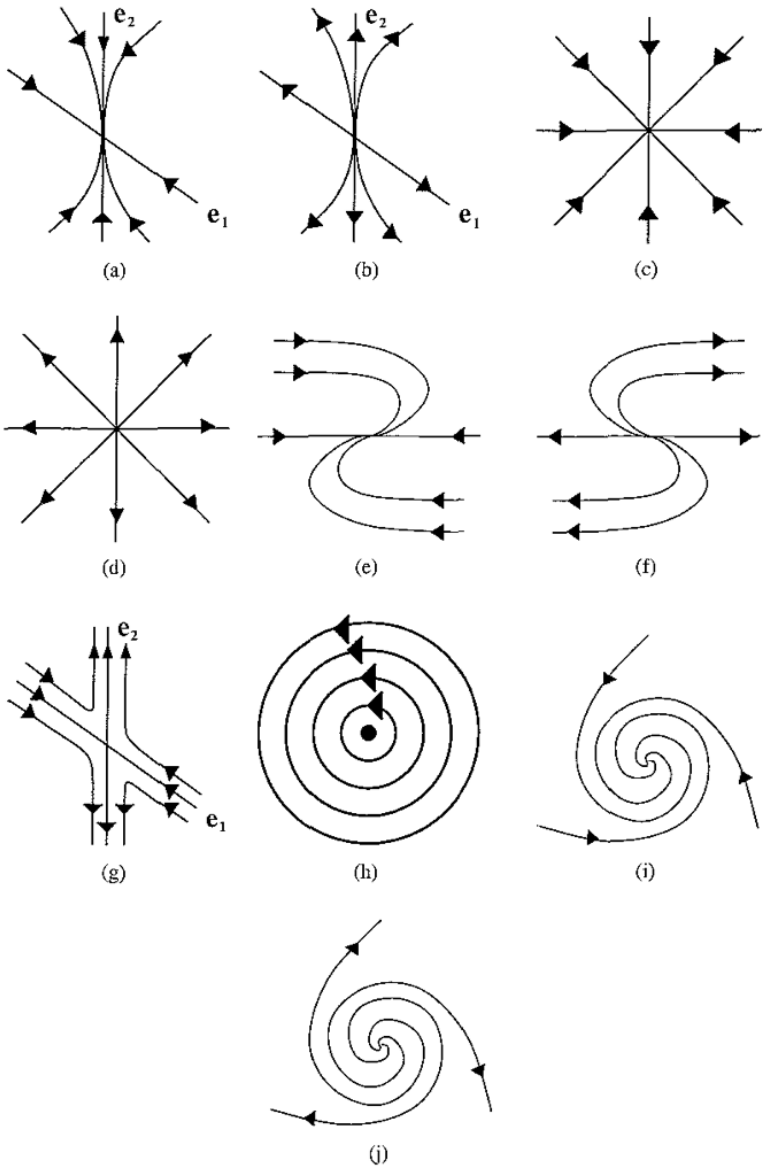


Fig 2 4. Phase portraits in the vicinity of (a) a stable node, (b) an unstable node, (c) a stable star, (d) an unstable star, (e) a stable improper node, (f) an unstable improper node, (g) a saddle point, (h) a centre, (i) a stable focus and (j) an unstable focus. Where marked, $\boldsymbol{e_1}$ and $\boldsymbol{e_2}$ are the eigendirections corresponding to eigenvalues λ_1 and λ_2, respectively.

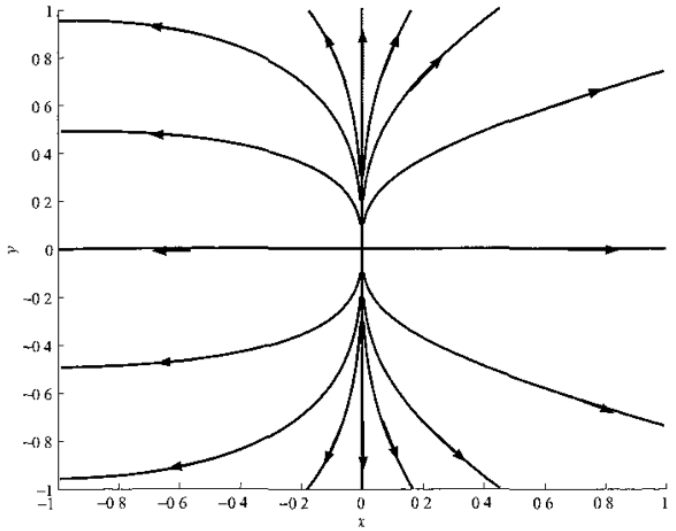


Fig 2.5 Phase portrait in a neighbourhood of the fixed point at the origin in Example 2.3

A stationary point, $\boldsymbol{x}_0$, is said to be **hyperbolic** if and only if the Jacobian, $Df\big|_{\boldsymbol{x}=\boldsymbol{x}_0}$, has no zero or purely imaginary eigenvalues. The point $\boldsymbol{x}_0$ is called a **sink** if all the eigenvalues have strictly negative real part, a **source** if all the eigenvalues have strictly positive real part, and a **saddle** otherwise.

Hyperbolic stationary points have the nice property that they persist under small perturbations of the governing differential equation. This means that if equation (2.1) is changed slightly to

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = f(\boldsymbol{x}) + \epsilon p(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^n \tag{2.16}$$

where $p(\boldsymbol{x})$ is a smooth vector field in $\mathbb{R}^n$, and $0 < \epsilon \ll 1$ is sufficiently small, then for each hyperbolic stationary point of equation (2.1) there is a hyperbolic stationary point of equation (2.16) that lies very close to the original one in phase space and is of the same stability type (sink, source or saddle).

We can see how this works by considering a hyperbolic fixed point at $\boldsymbol{x} = \boldsymbol{x}_0$ of equation (2.1). Fixed points of equation (2.16) satisfy

$$f(\boldsymbol{x}) + \epsilon p(\boldsymbol{x}) = \boldsymbol{0} \tag{2.17}$$

Now expanding this equation in a Taylor series around the original fixed point $\boldsymbol{x}_0$ gives

$$f(\boldsymbol{x}_0) + \epsilon p(\boldsymbol{x}_0) + \left(Df\big|_{\boldsymbol{x}=\boldsymbol{x}_0} + \epsilon Dp\big|_{\boldsymbol{x}=\boldsymbol{x}_0}\right)(\boldsymbol{x} - \boldsymbol{x}_0) + O(|\boldsymbol{x} - \boldsymbol{x}_0|^2) = \boldsymbol{0}. \tag{2.18}$$

We know that $f(\boldsymbol{x}_0) = \mathbf{0}$ because $\boldsymbol{x}_0$ is a fixed point of the original equation, and so from equation (2.18) we can deduce that

$$\boldsymbol{x} = \boldsymbol{x}_0 - \epsilon \left(Df\big|_{\boldsymbol{x}=\boldsymbol{x}_0} + \epsilon Dp\big|_{\boldsymbol{x}=\boldsymbol{x}_0} \right)^{-1} p(\boldsymbol{x}_0) + O(\epsilon^2) \equiv \boldsymbol{x}_1(\epsilon), \qquad (2.19)$$

and we have found a fixed point of the perturbed equation (2.16) at an $O(\epsilon)$ distance from the original fixed point of equation (2.1). Of course, for this to work the matrix $(Df\big|_{\boldsymbol{x}=\boldsymbol{x}_0} + \epsilon Dp\big|_{\boldsymbol{x}=\boldsymbol{x}_0})$ must have an inverse, and for that to be true it must have no zero eigenvalues. We know that $Df\big|_{\boldsymbol{x}=\boldsymbol{x}_0}$ has no zero eigenvalues, because the original fixed point $\boldsymbol{x}_0$ of equation (2.1) is hyperbolic. By continuity in ϵ, if we take ϵ small enough the eigenvalues of $(Df\big|_{\boldsymbol{x}=\boldsymbol{x}_0} + \epsilon Dp\big|_{\boldsymbol{x}=\boldsymbol{x}_0})$ will be bounded away from zero too, so the matrix will be invertible and our equation (2.19) for the perturbed fixed point will hold.

To show that the perturbed fixed point is of the same type as the original one, we have to show that none of the eigenvalues of $(Df + \epsilon Dp)$ can cross the imaginary axis as we move from $\boldsymbol{x} = \boldsymbol{x}_0$ when $\epsilon = 0$ (the original fixed point) to $\boldsymbol{x} = \boldsymbol{x}_1$ with $0 < \epsilon \ll 1$ (the perturbed fixed point). The original fixed point is hyperbolic, and so the real parts of its eigenvalues are bounded away from zero. By continuity in ϵ the real parts of the eigenvalues of $(Df\big|_{\boldsymbol{x}=\boldsymbol{x}_1(\epsilon)} + \epsilon Dp\big|_{\boldsymbol{x}=\boldsymbol{x}_1(\epsilon)})$ will also be bounded away from zero for small enough ϵ, and so cannot cross the imaginary axis. This means that the perturbed fixed point is also hyperbolic and of the same type as the original.

2.2 Local bifurcations from stationary points

At a bifurcation there is a sudden qualitative change in the flow in response to changes in one or more parameters in equation (2.1). There will be obvious differences in the phase portrait, and typically the number and stability properties of fixed points or periodic orbits will change. The parameters that lead to these changes are called **bifurcation parameters** and the point in parameter space at which the changes occur is called the **bifurcation point**. If we let $\boldsymbol{\mu}$ be a vector of bifurcation parameters, then $\boldsymbol{\mu} = \boldsymbol{\mu}_c$ for example, is the bifurcation point. Local bifurcation theory is concerned with changes in the flow in the neighbourhood of a fixed point or periodic orbit. There are also global bifurcations which affect the large-scale properties of the flow: we shall not consider them here, but if you are interested you will find a nice introductory account in Glendinning (1994).

In the remainder of this chapter we will look exclusively at local bifurcations from stationary points. The first thing to point out is that these bifurcations can only happen at parameter values for which the stationary point is nonhyperbolic. Changing parameter values a little bit is equivalent to making a small perturbation to the

governing ordinary differential equation, and we know that hyperbolic stationary points persist under those circumstances: the perturbed equation has a hyperbolic stationary point of the same stability type. So changes in the number or stability-type of fixed points can only happen when stationary points are nonhyperbolic. Typically the eigenvalues of the Jacobian, $Df|_{x=x_0}$, at the stationary point x_0 will depend upon the vector of bifurcation parameters, μ: to identify the bifurcation point you simply find the value or values of μ for which one or more of the eigenvalues is zero or purely imaginary.

Example 2.4 *Consider the equations*

$$\frac{dx}{dt} = \mu - x^2, \tag{2.20}$$

$$\frac{dy}{dt} = -y, \tag{2.21}$$

where $\mu, x, y \in \mathbb{R}$. *For* $\mu < 0$ *there are no fixed points, but for* $\mu > 0$ *there are stationary points at* $(\pm\sqrt{\mu}, 0)$, *where the Jacobian matrix is* $\begin{pmatrix} \mp 2\sqrt{\mu} & 0 \\ 0 & -1 \end{pmatrix}$, *respectively. The eigenvalues are thus* -1 *and* $\mp 2\sqrt{\mu}$ *at* $(\pm\sqrt{\mu}, 0)$, *respectively. So the fixed point* $(\sqrt{\mu}, 0)$ *is a stable node and* $(-\sqrt{\mu}, 0)$ *is a saddle. Both these fixed points are hyperbolic, except at* $\mu = 0$. *As* μ *passes through zero from negative to positive, there is a change in the flow, from there being no fixed points to there being two fixed points, one stable and one unstable: there is a qualitative change in the phase portrait. At* $\mu = 0$ *the points* $(\pm\sqrt{\mu}, 0)$ *coalesce into one nonhyperbolic fixed point at the origin. This is a saddle-node bifurcation: the phase portraits shown in Figure 2.6 illustrate the birth of a saddle and a node at* $\mu = 0$, *whence the name.*

2.2.1 Reduction to a centre manifold

To understand local bifurcations of stationary points we need a method of analysing the flow near a nonhyperbolic stationary point. In general we will want to investigate the case when a stable stationary point first loses stability, so the Jacobian will have eigenvalues with negative real part, corresponding to stable directions, and one or more eigenvalues with zero real part corresponding to the directions in which the the stationary point is just losing stability. We will not allow any eigenvalues to have positive real part as this would mean the stationary point already had unstable directions.

Let us write the system as

$$\frac{dx}{dt} = f(x, \mu), \quad x \in \mathbb{R}^n, \ \mu \in \mathbb{R}^m, \tag{2.22}$$

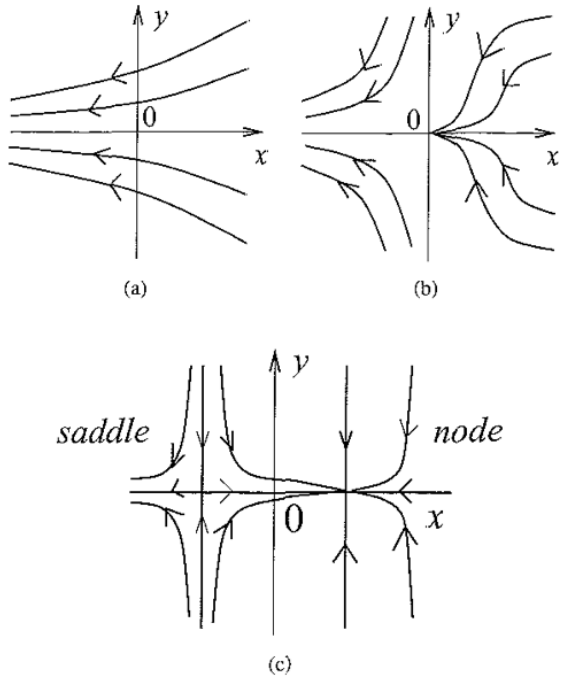


Fig 2 6. Phase portraits for the cases (a) $\mu < 0$, (b) $\mu = 0$ and (c) $\mu > 0$ of Example 2 4

where the nonhyperbolic stationary point is at $\boldsymbol{x} = \mathbf{0}$ It is helpful to separate out the linear and nonlinear parts of the equation so that

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = A\boldsymbol{x} + g(\boldsymbol{x}), \tag{2 23}$$

where $|g(\boldsymbol{x})| = O(|\boldsymbol{x}|^2)$ as $|\boldsymbol{x}| \to 0$. The matrix A is the Jacobian, $Df\big|_{(\mathbf{0},\mu)}$, and its eigenvalues will depend on μ The nonlinear function $g(\boldsymbol{x})$ may also depend on μ

Now we separate out the linearly stable directions, corresponding to eigenvalues with negative real part, from the directions corresponding to eigenvalues with zero real part by rewriting the system in the form

$$\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t} = B\boldsymbol{y} + g_y(\boldsymbol{y}, \boldsymbol{z}), \tag{2.24}$$

$$\frac{\mathrm{d}\boldsymbol{z}}{\mathrm{d}t} = C\boldsymbol{z} + g_z(\boldsymbol{y}, \boldsymbol{z}), \tag{2.25}$$

where all the eigenvalues of the matrix B have zero real part and all the eigenvalues of the matrix C have negative real part, and where $g_y(\boldsymbol{y}, \boldsymbol{z})$ and $g_z(\boldsymbol{y}, \boldsymbol{z})$ are the $\boldsymbol{y}$ and $\boldsymbol{z}$ components of $g(\boldsymbol{x})$ respectively. Sufficently close to $\boldsymbol{x} = \mathbf{0}$, the dynamics

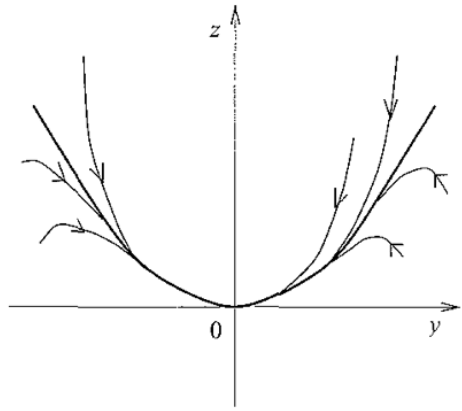


Fig 2 7. Trajectories are rapidly attracted towards the centre manifold (bold curve), where the subsequent dynamics unfold Arrows show the direction in which the trajectories are followed as time progresses

is dominated by the linear terms and $|z|$ decays towards zero exponentially fast, because the eigenvalues in the z directions have negative real part. If the dz/dt equation were purely linear ($g_z(y, z) \equiv \mathbf{0}$), we would have $|z| \to 0$, but in fact the Cz term that causes $|z|$ to decay will eventually be balanced by the nonlinear term $g_z(y, z)$, and z ends up on some manifold, $z = h(y)$, where $|h(y)|$ is no bigger than $O(|y|^2)$ as $|y| \to 0$ After the transients have died down, the dynamics is driven by y, which evolves much more slowly than z, because the eigenvalues in the y directions have zero real part, and so the growth or decay of $|y|$ is determined by terms no bigger than $O(|y|^2)$ in equation (2.24) This evolution takes place on the **centre manifold**, $z = h(y)$, illustrated in Figure 2.7, which can be determined from equations (2 24) and (2 25)

The **centre manifold theorem** guarantees that there exists a centre manifold passing through the origin and tangential to $z = \mathbf{0}$ at $y = \mathbf{0}$ It is invariant: in other words a trajectory that starts on the centre manifold stays on it The centre manifold is not necessarily unique, however. We shall move on to look at the use of the centre manifold theorem; if you are interested in details of its statement and proof you can find them in Carr (1981) One important comment is that systems that have some eigenvalues with positive real part can still have a centre manifold, even though we have chosen not to pursue that possibility. The centre manifold is still locally tangent to the eigenspace corresponding to eigenvalues with zero real part

Locally, the centre manifold can be written

$$z = h(y), \tag{2 26}$$

where $h(\mathbf{0}) = \mathbf{0}$, since it passes through the origin, and where $Dh\big|_{y=\mathbf{0}} = 0$, since it is tangential to the space $z = \mathbf{0}$ at $y = \mathbf{0}$

The centre manifold is invariant, so it must satisfy the equation

$$\frac{\mathrm{d}z}{\mathrm{d}t} = Dh\frac{\mathrm{d}y}{\mathrm{d}t}, \tag{2.27}$$

which after substituting from equations (2.24), (2.25) and (2.26) becomes

$$Ch(y) + g_z(y, h(y)) = Dh[By + g_y(y, h(y))] \tag{2.28}$$

To find the centre manifold, we need to solve equation (2.28) for $h(y)$. According to the centre manifold theorem there must be a solution near $y = \mathbf{0}$ We assume there is a polynomial approximation to $h(y)$ near $y = \mathbf{0}$: it is in general straightforward to find by expanding $h(y)$ in powers of y and equating the coefficients of powers of y on either side of equation (2.28)

After the transients have died away, the evolution of the flow on the centre manifold is governed by the equation

$$\frac{\mathrm{d}y}{\mathrm{d}t} = By + g_y(y, h(y)), \tag{2.29}$$

where we have simply substituted $z = h(y)$ into equation (2.24).

This procedure has reduced the original system of n equations to one of much lower order, for example m equations if $y \in \mathbb{R}^m$. (Note that m is the dimension of the centre manifold.) This is a very good thing, as the smaller the number of equations we have to deal with, the more likely we are to understand the system.

The centre manifold reduction can be carried out even as $n \to \infty$, provided that the eigenvalues, λ_i, of C satisfy

$$Re(\lambda_i) \leq -\delta < 0, \tag{2.30}$$

where $\delta > 0$ does not depend on n. In many of the situations we will examine in this book the centre manifold reduction will leave us with only one or a handful of equations to analyse in a system that started with an infinite number: this is very useful indeed.

2.2.2 *The extended centre manifold*

The centre manifold reduction is all very well, but what we would really like to be able to do is to analyse the behaviour near the stationary point as we vary the vector of bifurcation parameters μ through the bifurcation point. That means that the real part of one or more eigenvalues of A will pass from negative to positive as we pass

through the bifurcation point $\mu = \mu_c$, where the real parts are zero At first sight it looks as if we can't use a centre manifold reduction, because the fact that the eigenvalues change sign as μ varies means we can't separate them into those with negative and those with zero real part We use a trick to solve this problem. First the vector of bifurcation parameters, μ, is chosen so that the bifurcation point is at $\mu = \mu_c = \mathbf{0}$. Then we declare that μ is now a variable vector just like y and z, and tack the equation

$$\frac{d\mu}{dt} = \mathbf{0} \tag{2.31}$$

onto to our system So μ is a variable vector – it's just that it doesn't evolve in a very interesting way. This might seem rather pointless, but magically the full set of governing equations

$$\frac{d\mu}{dt} = \mathbf{0} \tag{2.32}$$

$$\frac{dy}{dt} = B'y + g_y'(y, z, \mu) \equiv By + g_y(y, z), \tag{2.33}$$

$$\frac{dz}{dt} = C'z + g_z'(y, z, \mu) \equiv Cy + g_z(y, z), \tag{2.34}$$

now satisfies the conditions for a centre manifold reduction: there is a nonhyperbolic stationary point at the origin, $\mu = y = z = \mathbf{0}$, and all the linear eigenvalues of B' and C' have either zero or negative real part Note that because μ is on an equal footing with y and z in this formulation, the entries in B and C that depend on μ now give rise to nonlinear terms that must be included in g_y' and g_z' respectively, while the new matrices B' and C' must be independent of μ The primes ($'$) indicate that the linear and nonlinear terms have been redefined to account for this shifting of the μ-dependent terms out of the matrices and into the nonlinear functions What we have done is extended the system into parameter space: we will now be able to find an extended centre manifold on which we will be able to examine the flow near the stationary point for parameter values close to the bifurcation point on either side.

We find the **extended centre manifold**, $z = h(y, \mu)$, by expanding in powers of y and μ. Eliminating z from equation (2.33) then gives

$$\frac{dy}{dt} = By + g_y(y, h(y, \mu)), \tag{2.35}$$

and we know that μ is a constant from integration of equation (2.32). Equation (2.35) describes the evolution on the extended centre manifold close to the bifurcation point

Example 2.5 *Consider the system*

$$\frac{dx}{dt} = \mu x - xy, \tag{2.36}$$

$$\frac{dy}{dt} = -y + x^2, \tag{2.37}$$

where $\mu, x, y \in \mathbb{R}$. *There is a stationary point at the origin,* $x = y = 0$, *for all values of* μ, *and the Jacobian there has eigenvalues* μ *and* -1. *At* $\mu = 0$ *the stationary point is nonhyperbolic and there is a bifurcation.*

To analyse behaviour near the stationary point for values of μ *close to zero, we add the equation*

$$\frac{d\mu}{dt} = 0 \tag{2.38}$$

and look for the extended centre manifold. Clearly μ *has a growth rate eigenvalue of zero. Now that* μ *is a variable, the linear part of equation (2.36) is zero, so* x *also has a zero eigenvalue. The linear part of equation (2.37) gives an eigenvalue of* -1 *for* y. *Since* y *is the linearly decaying variable, and* μ *and* x *are linearly stationary, we look for the extended centre manifold in the form* $y = h(x, \mu)$. *The invariance of the centre manifold gives us the equation*

$$\frac{dy}{dt} = \frac{\partial h}{\partial x}\frac{dx}{dt} + \frac{\partial h}{\partial \mu}\frac{d\mu}{dt}. \tag{2.39}$$

Substituting from equations (2.36) and (2.38), along with $y = h(x, \mu)$, *leads to*

$$-h(x, \mu) + x^2 = \frac{\partial h}{\partial x}(\mu x - xh(x, \mu)). \tag{2.40}$$

Now we expand $h(x, \mu)$ *as a polynomial in* x *and* μ, *with* h *at least as small as quadratic near* $x = \mu = 0$ *(so that the centre manifold is tangent to the space* $y = 0$ *at the origin) to get*

$$h = ax^2 + bx\mu + c\mu^2 + \cdots \tag{2.41}$$

where a, b *and* c *are constants and the dots remind us that there will be higher-order terms as well. Substituting this into equation (2.40), gives*

$$x^2 - ax^2 - bx\mu - c\mu^2 - \cdots = (2ax + b\mu)x(\mu - ax^2 - bx\mu - c\mu^2 - \cdots). \tag{2.42}$$

This equation must hold for all values of x *and* μ, *so the coefficients of each term* $x^i\mu^j$, *for* $i, j \in \mathbb{Z}$, *on the left- and righthand sides must be equal. Equating the coefficients of* x^2 *on either side of equation (2.42) gives* $1 - a = 0$, *or* $a = 1$. *Equating the coefficients of* $x\mu$ *gives* $b = 0$ *and equating the coefficients of* μ^2

gives $c = 0$. *So to leading order the extended centre manifold is*

$$y = x^2, \tag{2.43}$$

and the flow on the extended manifold evolves according to the equation

$$\frac{dx}{dt} = \mu x - x^3, \tag{2.44}$$

where μ *is a constant.*

We can also derive equation (2.44) using the method of **adiabatic elimination**: assume that μ is small, for example $\mu \sim \epsilon^2$ with $|\epsilon| \ll 1$. Now, we know that on the extended centre manifold, the evolution is driven by x, so we must balance all the terms in the equation for dx/dt, namely

$$\frac{dx}{dt} = \mu x - xy, \tag{2.45}$$

to ensure that they are all of the same order in ϵ: we find $t \sim \epsilon^{-2}$, so that $d/dt \sim \epsilon^2$, and $y \sim \epsilon^2$. Now turning to equation (2.37), we see that the dy/dt term is of order ϵ^4, but the $-y$ term is of order ϵ^2. In order to balance this last term, we must have $x \sim \epsilon$. The term dy/dt will later be balanced by terms coming from a higher-order expansion of y in terms of x and μ: the fact that this term is small reflects the fact that the y coordinate changes only slowly once the trajectory has reached the centre manifold. The equations look like this

$$\frac{dx}{dt} = \mu x - xy, \tag{2.46}$$

$$O(\epsilon^3) \quad O(\epsilon^3)\ O(\epsilon^3)$$

$$\frac{dy}{dt} = -y + x^2. \tag{2.47}$$

$$O(\epsilon^4) \quad O(\epsilon^2)\ O(\epsilon^2)$$

Equating terms at $O(\epsilon^2)$ in equation (2.47), we find $y = x^2$, as before. The y coordinate is determined directly from the x coordinate: we say that it is **slaved** to the x coordinate. Now we can substitute the expression for y into equation (2.46) to get the equation

$$\frac{dx}{dt} = \mu x - x^3, \tag{2.48}$$

which describes the dynamics on the extended centre manifold. This simple method only gives the leading-order approximation to the centre manifold.

2.3 Normal forms for bifurcations

Let us return to the equation

$$\frac{dx}{dt} = f(x, \mu) \quad x \in \mathbb{R}^n, \ \mu \in \mathbb{R}^m, \tag{2.49}$$

where f is nonlinear, μ is a vector of bifurcation parameters and there is a stationary point at $x = 0$ **Poincaré's linearization theorem** says that under certain conditions on the eigenvalues of $Df|_{(0\ \mu)}$ there is a nonlinear near-identity transformation

$$y = x + g(x), \quad y \in \mathbb{R}^n, \tag{2.50}$$

in the neighbourhood of the origin, which is invertible for sufficiently small $|x|$, such that the evolution equation for y is linear and takes the form

$$\frac{dy}{dt} = Df|_{(0\ \mu)} y \tag{2.51}$$

For further details of this theorem look in Glendinning (1994) or Arnol'd (1983). If we could make this transformation for every nonlinear system for every value of μ, then lots of problems would be much easier to solve. However, it doesn't always work, because the conditions for Poincaré's linearization theorem to hold are quite strong In particular, the result does not hold at a local bifurcation point. If we attempt to linearise the system there, we find that we cannot get rid of some of the nonlinear terms. The transformed system now consists of the nonlinear terms we are stuck with. If we add back in the linear terms that vanish exactly at the bifurcation point – a process known as **unfolding** – the result is known as the **normal form** of the bifurcation; it contains all the essential information about the character of the bifurcation. Normal forms are usually written down using simpler, less formal methods

2.4 Codimension-one bifurcations

The **codimension** of a bifurcation is the difference between the dimension of the bifurcation-parameter space and the dimension of the object (for example surface, line or point) that gives the location of the bifurcation in that space For example, if a bifurcation occurs on the (one-dimensional) line $\mu_1 = 0$ in a two-dimensional parameter space, $\mu = (\mu_1, \mu_2)$, then the codimension of the bifurcation is $2 - 1 = 1$. Another way of looking at this is that the codimension is equal to the number of bifurcation parameters that you need to vary in order to reach the locus of the bifurcation from a typical point in parameter space. Starting from a typical point (μ_1, μ_2) in the example just given, you only need to vary μ_1 in order

to reach $\mu_1 = 0$, so the bifurcation has codimension one. Of course, if you start at a special point where $\mu_1 = 0$ there is no need to vary any parameters to reach the bifurcation, because you are there already. The mention of 'typical points' is intended to exclude this situation.

We are going to take a quick look at codimension-one bifurcations in this section as they are the simplest examples to understand, and turn out to be important in pattern formation. For these bifurcations there is only one bifurcation parameter, $\mu \in \mathbb{R}$, to vary in order to find the bifurcation point. We will assume that when $\mu = 0$, there is a nonhyperbolic stationary point at $\boldsymbol{x} = \mathbf{0}$, in other words that there is a bifurcation at $\mu = 0$, and we will look at the local behaviour close to $(\boldsymbol{x}, \mu) = (\mathbf{0}, 0)$.

2.4.1 Stationary bifurcations

We will first consider bifurcations that correspond to a real eigenvalue passing through zero. This type of bifurcation is known as **steady** or **stationary**.

Saddle-node bifurcation

We have already met the saddle-node bifurcation in Example 2.4; its normal form is given by the equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mu - ax^2 + \qquad , \tag{2.52}$$

where $x, \mu \in \mathbb{R}$ and a is a real constant. The bifurcation parameter is μ. As we saw in Example 2.4 for $a = 1$, there are fixed points at $x = \pm\sqrt{\mu/a}$ for $\mu/a > 0$, and no fixed points for $\mu/a < 0$, with the bifurcation point at $\mu = 0$. The growth rate eigenvalue for perturbations to the fixed point at $\pm\sqrt{\mu/a}$ is

$$\left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x=\pm\sqrt{\mu/a}} = -2ax|_{x=\pm\sqrt{\mu/a}} = \mp 2a\sqrt{\mu/a}, \tag{2.53}$$

respectively, where $f(x) = \mu - ax^2$ is the righthand side of the normal form equation. So for $a > 0$ the fixed point at $x = +\sqrt{\mu/a}$ is stable and the one at $x = -\sqrt{\mu/a}$ is unstable, while for $a < 0$ the stabilities are interchanged. Note that at the bifurcation point the growth rate eigenvalue is zero, as expected, and there is a nonhyperbolic fixed point at the origin. We can summarise this information in the **bifurcation diagram** shown in Figure 2.8, which plots the amplitude x of the fixed points as a function of the bifurcation parameter μ. The figure shows the case $a > 0$. Stable fixed points are shown as solid lines and unstable ones as dotted lines.

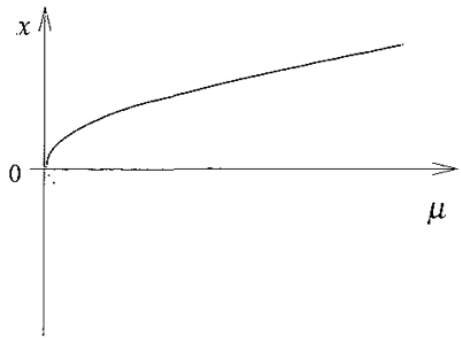


Fig 2 8. Bifurcation diagram for the saddle-node bifurcation. Stable fixed points are shown as bold lines, and unstable fixed points as dotted lines

The saddle-node bifurcation is the generic codimension-one stationary bifurcation: it is the bifurcation you would expect to find when no special conditions have been imposed on the system.

To make the notion of the **genericity** of a bifurcation a bit clearer we can say that the bifurcation problem

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = f(\boldsymbol{x}, \mu) \tag{2.54}$$

is **generic** for its class, characterised by constraints on $f(\boldsymbol{x}, \mu)$, if for sufficiently small $\epsilon > 0$ the perturbed problem

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = f(\boldsymbol{x}, \mu) + \epsilon v(\boldsymbol{x}, \mu) \tag{2.55}$$

has the same type of bifurcation at a nearby value of μ for all perturbations $v(\boldsymbol{x}, \mu)$ such that it remains in the same class. This is a bit of a hand-waving definition of genericity, but it is adequate for our purposes

In the case of the saddle-node bifurcation the only constraint on $f(x, \mu)$ is that it be a smooth function of $x, \mu \in \mathbb{R}$. Consider the smooth perturbation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mu - ax^2 + \epsilon v(x, \mu) + \cdots \tag{2.56}$$

of the original bifurcation problem, where without loss of generality, we consider the case $a > 0$ Looking close to the fixed point at $\mu = 0$, $x = 0$, the perturbation takes the form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mu(1 + \epsilon v_1 + \epsilon v_2 \mu) + \epsilon(v_3 + v_4\mu)x - (a + \epsilon v_5)x^2 + \cdots \tag{2.57}$$

up to quadratic order in x and μ, for some constants v_i, $i = 1, \ldots, 5$. The stationary solutions to this equation,

$$x = \frac{1}{2(a+\epsilon v_5)}\Big\{\epsilon(v_3+v_4\mu) \pm\sqrt{\epsilon^2(v_3+v_4\mu)^2+4\mu(a+\epsilon v_5)(1+\epsilon v_1+\epsilon v_2\mu)}\Big\}, \tag{2.58}$$

exist only for

$$\epsilon^2(v_3+v_4\mu)^2+4\mu(a+\epsilon v_5)(1+\epsilon v_1+\epsilon v_2\mu) > 0, \tag{2.59}$$

or equivalently,

$$\mu > \mu_c \equiv -\frac{\epsilon^2 v_3^2}{4a} + O(\epsilon^3) \tag{2.60}$$

There is a saddle-node bifurcation at $\mu = \mu_c$. Since $\mu_c \sim \epsilon^2$, the bifurcation point is close to that of the original system at $\mu = 0$. Thus the saddle-node bifurcation is generic for one-dimensional steady-state bifurcation problems.

We will now go on to look at stationary bifurcations that occur when the system is special in some way, so that there are further constraints on $f(x, \mu)$.

Transcritical bifurcation

The simplest codimension-one stationary bifurcation with a fixed point that persists for all values of the bifurcation parameter is the **transcritical** bifurcation. Its normal form is

$$\frac{dx}{dt} = \mu x + ax^2 + \cdots, \tag{2.61}$$

where $x, \mu \in \mathbb{R}$ and a is a real constant. This has fixed points at $x = 0$ and $x = -\mu/a$ for all values of μ. Linearising around the fixed point at $x = 0$ by setting $x = 0 + \eta(t)$, where $|\eta| \ll 1$ we see that

$$\frac{d\eta}{dt} = \mu\eta, \tag{2.62}$$

so the growth rate eigenvalue is μ and the point $x = 0$ is stable for $\mu < 0$ and unstable for $\mu > 0$. Linearising around the second fixed point at $x = -\mu/a$ by setting $x = -\mu/a + \eta(t)$ with $|\eta| \ll |\mu/a|$ gives

$$\frac{d\eta}{dt} = -\mu\eta \tag{2.63}$$

The eigenvalue is $-\mu$ and so the point $x = -\mu/a$ is unstable for $\mu < 0$ and stable for $\mu > 0$. The fixed points and their stability are shown in the bifurcation diagram

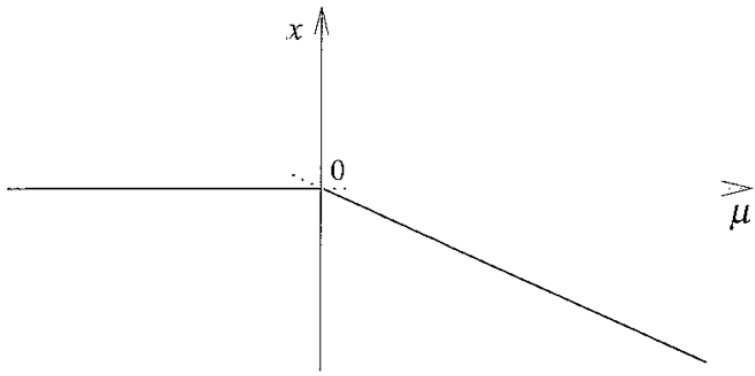


Fig 2 9. Bifurcation diagram for the transcritical bifurcation Stable fixed points are shown as bold lines, and unstable fixed points as dotted lines.

in Figure 2 9. At $\mu = 0$, there is a bifurcation point, where the stability of the two fixed points is exchanged.

The additional constraint on $f(x, \mu)$ for a transcritical bifurcation beyond that for a saddle-node is

$$f(0, \mu) = 0, \quad \forall \mu \in \mathbb{R}, \tag{2.64}$$

which forces $x = 0$ to be a fixed point of equation (2.54) for all μ It is easy to check that within the class of stationary bifurcation problems satisfying equation (2.64) transcritical bifurcations are generic. On the other hand they are not generic within the wider class of stationary bifurcation problems in one dimension. Choosing the following smooth perturbation that does not satisfy (2.64)

$$\frac{dx}{dt} = \epsilon v_1 + \mu x + ax^2 + \cdots, \tag{2.65}$$

we find stationary solutions at

$$x = \frac{1}{2a}\left(-\mu \pm \sqrt{\mu^2 - 4\epsilon v_1 a}\right), \tag{2.66}$$

as long as $\mu^2 > 4\epsilon v_1 a$. If $av_1 < 0$ there are stationary solutions for all values of μ, and there is no bifurcation. On the other hand if $av_1 > 0$ stationary solutions only exist in the regions $x < -2\sqrt{\epsilon v_1 a}$ and $x > 2\sqrt{\epsilon v_1 a}$, and the transcritical bifurcation breaks up into two saddle-nodes The two situations are illustrated in Figure 2.10. It is easy to check that the stabilities of the various solution branches are as given in the diagrams. It is no surprise that if we see bifurcations in the perturbed case they are saddle-nodes, because these are generic when the constraint $f(0, \mu) = 0$ is broken

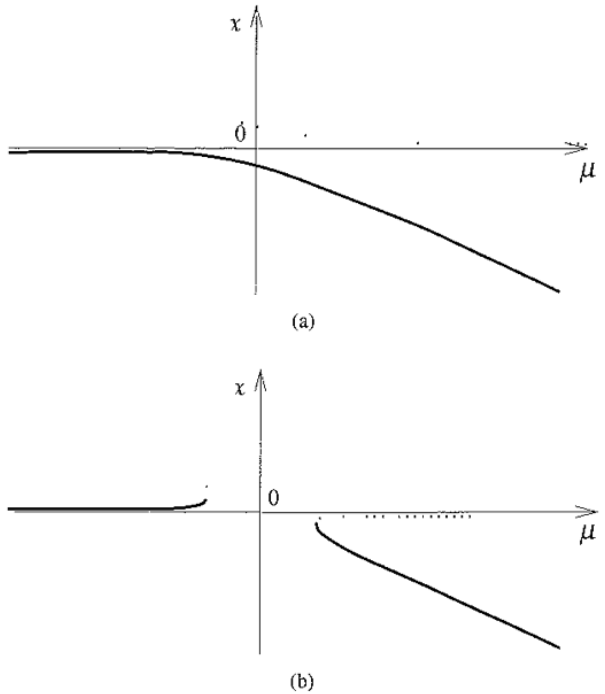


Fig 2.10. Bifurcation diagrams for the perturbed transcritical bifurcation described by equation (2.65) with $a > 0$ and (a) $v_1 < 0$ or (b) $v_1 > 0$. Stable fixed points are marked with bold lines and unstable fixed points with dotted lines.

Pitchfork bifurcation

If we require both that $x = 0$ is a fixed point for all values of the bifurcation parameter, and also that the system is symmetric under the transformation $x \to -x$, then we end up with a **pitchfork** bifurcation. There are two possible normal forms. The first case is known as the **subcritical** pitchfork bifurcation:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mu x + ax^3 + \cdots, \tag{2.67}$$

where $x, \mu \in \mathbb{R}$ and $a > 0$ is a real constant. The fixed points are $x = 0$ for all μ and $x = \pm\sqrt{-\mu/a}$ for $\mu < 0$. Linearising around $x = 0$ by setting $x = 0 + \eta(t)$ with $|\eta| \ll 1$ we see that

$$\frac{\mathrm{d}\eta}{\mathrm{d}t} = \mu\eta, \tag{2.68}$$

just as for the transcritical bifurcation, so again $x = 0$ is stable for $\mu < 0$ and unstable for $\mu > 0$. Now linearising around the other fixed points by setting

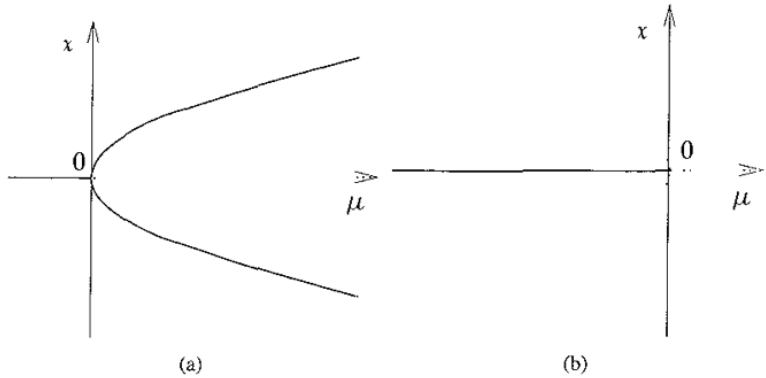


Fig 2.11. Bifurcation diagrams for the (a) supercritical and (b) subcritical pitchfork bifurcations. Stable fixed points are marked with bold lines and unstable fixed points with dotted lines.

$x = \pm\sqrt{-\mu/a} + \eta(t)$, with $|\eta| \ll \sqrt{-\mu/a}$ gives

$$\frac{\mathrm{d}\eta}{\mathrm{d}t} = -2\mu\eta, \tag{2.69}$$

so the solutions $x = \pm\sqrt{-\mu/a}$ are unstable for $\mu < 0$ (and do not exist for $\mu > 0$).

The second case is known as the **supercritical** pitchfork bifurcation:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mu x - ax^3 + \quad \cdot, \tag{2.70}$$

where $x, \mu \in \mathbb{R}$ and $a > 0$ is a real constant. The fixed points are $x = 0$ for all μ and $x = \pm\sqrt{\mu/a}$ for $\mu > 0$. Again the point $x = 0$ is stable for $\mu < 0$ and unstable for $\mu > 0$. Setting $x = \pm\sqrt{\mu/a} + \eta(t)$, with $|\eta| \ll \sqrt{\mu/a}$ and linearising in η gives

$$\frac{\mathrm{d}\eta}{\mathrm{d}t} = -2\mu\eta, \tag{2.71}$$

so the solutions $x = \pm\sqrt{\mu/a}$ are stable for $\mu > 0$ (and do not exist for $\mu < 0$).

In both cases there is a bifurcation at $\mu = 0$ where the zero solution loses stability. The pitchfork bifurcations are illustrated in Figure 2.11.

The defining characteristic of the subcritical bifurcation is that the fixed point solution branches bifurcating away from the zero solution exist in the region where the zero solution is stable, **before** the critical point at which the zero solution loses stability, whereas for a supercritical bifurcation the bifurcating branches exist in the region where the zero solution is unstable, in other words **after** the bifurcation point. For a given value of the bifurcation parameter, supercritical solution branches have more stable eigenvalues than the solution from which they have just

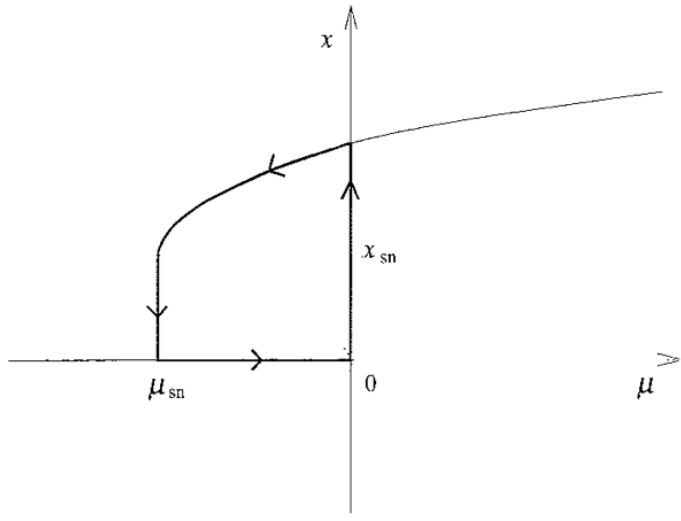


Fig 2 12. Bifurcation diagram showing the turnaround of a subcritical solution branch. There is a subcritical pitchfork bifurcation at $\mu = 0$ and a saddle-node bifurcation at $\mu = \mu_{sn}$ The hysteresis loop is followed in the direction of the arrows as μ is increased and decreased

bifurcated, whereas subcritical branches have fewer stable eigenvalues. In a one-dimensional system this means that subcritical branches are unstable and supercritical branches are stable At a supercritical bifurcation there is a smooth transition from the stable zero solution to the stable bifurcating branch as the bifurcation parameter μ increases through zero and the zero solution loses stability This is not the case for a subcritical bifurcation: as you increase the bifurcation parameter through zero the zero solution loses stability, but there is no small amplitude stable solution for $\mu > 0$ and so the amplitude x must jump from zero to some other stable solution at large amplitude not predicted by our local analysis. In general, one would expect a situation like the one illustrated in Figure 2.12: the unstable bifurcating branch gains stability in a saddle-node bifurcation at $x = x_{sn}$, $\mu = \mu_{sn}$. There would be a discontinuous jump from the zero solution to the high-amplitude stable solution at $\mu = 0$, and even in the region $\mu_{sn} < \mu < 0$ if the system were subjected to a large enough perturbation. On subsequently decreasing μ through zero, the high-amplitude solution would be stable until $\mu = \mu_{sn}$ at which point there would be a discontinuous jump down to the zero solution In the absence of large perturbations, the system would follow the **hysteresis loop** shown in Figure 2.12 as μ is first increased and then decreased We say that there is **hysteresis** in the region $\mu_{sn} < \mu < 0$, because the solution adopted by the system for a value of μ in the hysteretic range depends on the history of its evolution.

Pitchfork bifurcations are generic for the class of one-dimensional stationary bifurcation problems satisfying

$$f(0, \mu) = 0, \tag{2.72}$$

$$f(-x, \mu) = -f(x, \mu), \tag{2.73}$$

in other words those that not only have a stationary solution $x = 0$ for all μ, but are also symmetric under the reflection symmetry $x \to -x$. The reflection symmetry requires f to be odd in x so that

$$\frac{dx}{dt} = f(x, \mu) \tag{2.74}$$

is preserved under the transformation Letting $x \to -x$ gives

$$-\frac{dx}{dt} = f(-x, \mu) = -f(x, \mu), \tag{2.75}$$

which reduces to equation (2.74) on cancellation of the minus sign on both sides. Genericity of the pitchfork bifurcation under the constraints (2.72) and (2.73) is straightforward to check.

Pitchforks are not generic in the wider class of one-dimensional stationary bifurcation problems that satisfy only $f(0, \mu) = 0$ (and hence neither are they generic for the even wider class where f need only be smooth) It is easy to see this by perturbing the normal form to

$$\frac{dx}{dt} = \mu x + \epsilon v_1 x^2 - ax^3 + \cdots, \tag{2.76}$$

where v_1 is a constant, and where we shall consider the supercritical pitchfork with $a > 0$ (though a similar analysis is possible in the subcritical case) The stationary solutions are $x = 0$ and

$$x = \frac{1}{2a}\left(\epsilon v_1 \pm \sqrt{\epsilon^2 v_1^2 + 4a\mu}\right) \tag{2.77}$$

The nonzero solutions only exist for

$$\mu > \mu_c \equiv -\frac{\epsilon^2 v_1^2}{4a}, \tag{2.78}$$

and there is a saddle-node bifurcation at $\mu = \mu_c$ where this pair of solutions is formed There is also a transcritical bifurcation at $\mu = 0$. The bifurcation diagram is shown in Figure 2.13 As a result of the perturbation, the bifurcation problem (2.76) breaks the $x \to -x$ symmetry, but it still lies in the class where $f(0, \mu) = 0$, for which transcritical bifurcations are generic It is natural then that there is a transcritical, rather than a pitchfork bifurcation at the origin in the perturbed case.

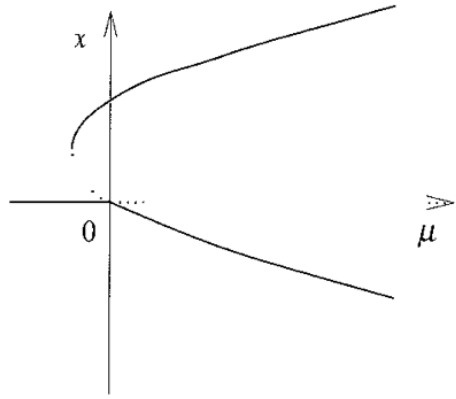


Fig. 2.13 Bifurcation diagram for the perturbed pitchfork described by equation (2.76) with $a > 0$ and $v_1 > 0$. Stable solutions are shown by bold lines and unstable solutions by dotted lines.

2.4.2 Hopf or oscillatory bifurcation

All the bifurcations we have looked at so far have been stationary, with a real eigenvalue passing through zero. We can also have a codimension-one bifurcation where the real part of a complex conjugate pair of eigenvalues passes through zero, the imaginary part remaining nonzero: this is a **Hopf** or **oscillatory** bifurcation. The normal form is given by

$$\frac{dx}{dt} = \mu x - \omega y + a(x^2 + y^2)x - b(x^2 + y^2)y, \tag{2.79}$$

$$\frac{dy}{dt} = \mu y + \omega x + a(x^2 + y^2)y + b(x^2 + y^2)x, \tag{2.80}$$

where $x, y, \mu \in \mathbb{R}$ and ω, a, b are real constants. The origin $x = y = 0$ is a fixed point of the equations for all parameter values. If we consider equations (2.79) and (2.80) to be of the form

$$\frac{d\boldsymbol{x}}{dt} = f(\boldsymbol{x}), \tag{2.81}$$

where $\boldsymbol{x} = \begin{pmatrix} x \\ y \end{pmatrix}$, then the Jacobian at the origin is given by

$$Df|_{\mathbf{0}} = \begin{pmatrix} \mu & -\omega \\ \omega & \mu \end{pmatrix}, \tag{2.82}$$

and its eigenvalues are $\mu \pm i\omega$. We expect a Hopf bifurcation when $\mu = 0$ and $\omega \neq 0$. (If $\omega = 0$ there is still a bifurcation, but it is not oscillatory.)

The easiest way to see what is going on is to go into polar coordinates setting $x = r\cos\theta$ and $y = r\sin\theta$, where $r > 0$ and $0 \leq \theta < 2\pi$. The equations for r and

θ are given by

$$\frac{dr}{dt} = \mu r + ar^3, \tag{2.83}$$

$$\frac{d\theta}{dt} = \omega + br^2. \tag{2.84}$$

The evolution of both r and θ depends only on the value of r, and not on θ, so we can determine the stability of the whole system by looking at the stability of solutions to perturbations in r. Equation (2.83) is the normal form for a pitchfork bifurcation in r, subcritical if $a > 0$ and supercritical if $a < 0$. So a Hopf bifurcation is a pitchfork in the r direction with some rotation in the θ direction thrown in: since ω is nonzero, θ is not steady close to the origin at the bifurcation point.

The stationary solutions for r are $r = 0$ for all μ, and $r = \sqrt{-\mu/a}$ for $\mu/a < 0$. Note that r is a radius, and so is non-negative. The point $r = 0$ is the fixed point at the origin, while at $r = \sqrt{-\mu/a}$ there is a periodic orbit, as long as $d\theta/dt$ is nonzero, which requires that $\omega - b\mu/a \neq 0$, so that the trajectory traces out the circle $r = \sqrt{-\mu/a}$ over and over again. Linearising around the solution $r = 0$ by setting $r = 0 + \eta(t)$, with $|\eta| \ll 1$, we find that

$$\frac{d\eta}{dt} = \mu\eta, \tag{2.85}$$

showing that $r = 0$ is stable for $\mu < 0$ and unstable for $\mu > 0$. Now setting $r = \sqrt{-\mu/a} + \eta(t)$, with $|\eta| \ll \sqrt{-\mu/a}$, and linearising in η gives

$$\frac{d\eta}{dt} = -2\mu\eta. \tag{2.86}$$

In the case $a > 0$, the solution $r = \sqrt{-\mu/a}$ exists for $\mu < 0$, and we see from equation (2.86) that it is then unstable. This is a subcritical Hopf bifurcation since the bifurcating solution, in this case a periodic orbit, exists in the region where the zero solution is stable. However, if $a < 0$ then the solution $r = \sqrt{-\mu/a}$ exists for $\mu > 0$ and from equation (2.86) is stable. This is a supercritical Hopf bifurcation since the bifurcating solution exists where the zero solution is unstable. This is directly analogous to case of the pitchfork bifurcation, because we do indeed have a pitchfork bifurcation in the r variable. The bifurcation point is at $\mu = 0$ where the zero solution, $r = 0$, loses stability. The bifurcation diagrams for sub- and supercritical Hopf bifurcations are illustrated in Figure 2.14.

For the case $\omega > \mu b/a$, where the periodic orbit is followed in an anticlockwise direction, the phase plane portraits for the sub- and supercritical Hopf bifurcations are shown in Figures 2.15 and 2.16 respectively.

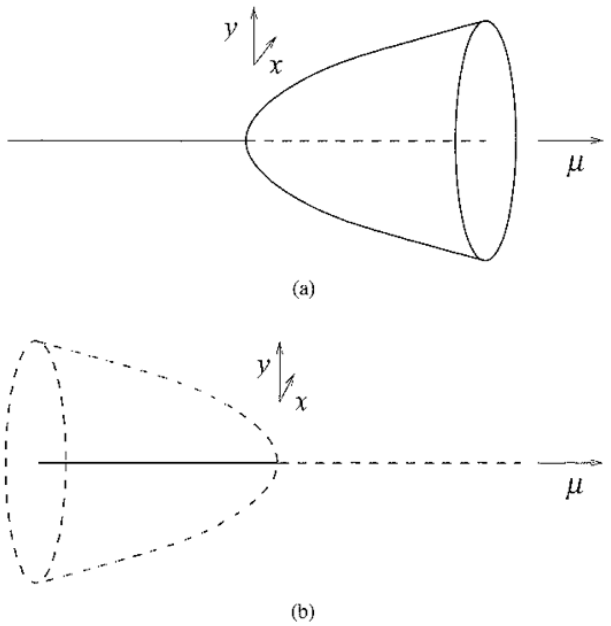


Fig 2.14 Bifurcation diagrams for the (a) supercritical and (b) subcritical Hopf bifurcations. Stable solutions are shown as bold lines, and unstable solutions as dotted lines. (The axes are drawn with the origin shifted.)

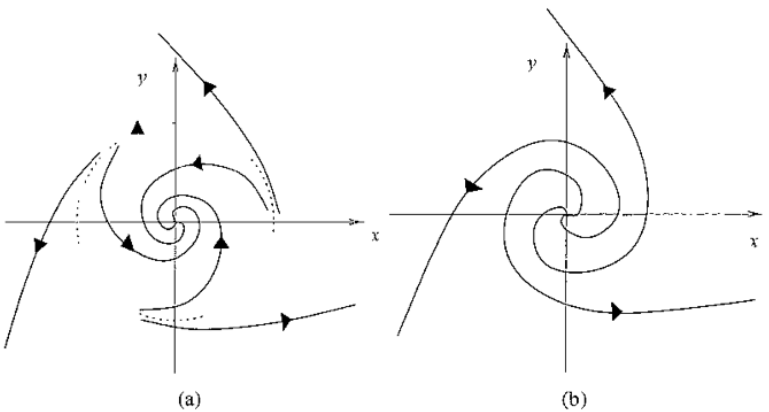


Fig 2.15 Phase plane diagrams for the subcritical Hopf bifurcation when (a) $\mu < 0$ and (b) $0 < \mu < a\omega/b$, for $\omega > 0$, $b > 0$ and $a > 0$. The limit cycle is shown as a dotted line because it is unstable.

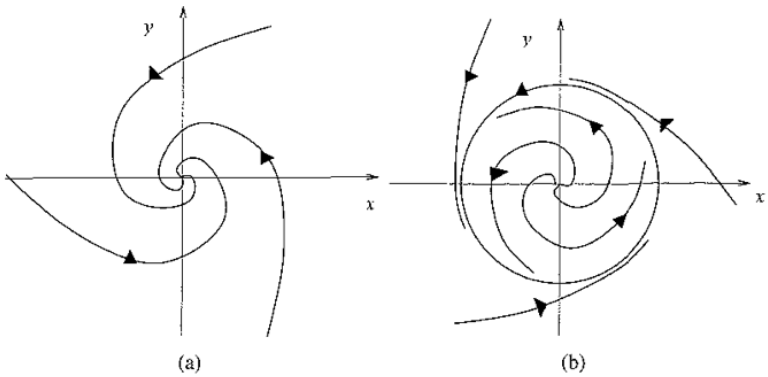


Fig 2.16 Phase plane diagrams for the supercritical Hopf bifurcation when (a) $a\omega/b < \mu < 0$ and (b) $\mu > 0$, for $\omega > 0$, $b > 0$ and $a < 0$.

Exercises

2.1 Identify the following bifurcations, put them in normal form, and sketch the bifurcation diagram:

(a) $$\frac{dx}{dt} = 3\mu + 27 + \mu x + 12x + x^2,$$

(b) $$\frac{dx}{dt} = -\mu - 98 - 20x - x^2,$$

(c) $$\frac{dx}{dt} = -3\mu + 24 - \mu x + 11x + x^2,$$

(d) $$\frac{dx}{dt} = \mu - 16 - 6x - x^2.$$

2.2 Reduce the following systems to the equation on the extended centre manifold (up to cubic order), identify the bifurcation there and sketch the bifurcation diagram:

(a) $$\frac{dx}{dt} = (\mu - 6)x + (\mu + 4)y + 2xy + 2y^2,$$
$$\frac{dy}{dt} = (\mu + 4)x + (\mu - 6)y - 2x^2 - 2xy.$$

(b) $$\frac{dx}{dt} = -\frac{1}{2}\mu(x + y) + y + \frac{5}{16}x^2 + \frac{1}{8}xy - \frac{3}{16}y^2,$$
$$\frac{dy}{dt} = -\frac{1}{2}\mu(x + y) + x + \frac{3}{16}x^2 - \frac{1}{8}xy - \frac{5}{16}y^2.$$

(c) $$\frac{dx}{dt} = y - x - x^2,$$
$$\frac{dy}{dt} = \mu x - y - y^2$$

2.3 By writing $dx/dt = y$ and finding dy/dt, or otherwise, sketch in the (μ, λ) plane the lines giving the locations of the stationary bifurcation and the oscillatory bifurcation

from the trivial solution $x = \mathrm{d}x/\mathrm{d}t = 0$ for the equation

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + \mu\frac{\mathrm{d}x}{\mathrm{d}t} + \lambda x - x^3 = 0$$

2.4 Consider the following equations in polar coordinates (r, θ):

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \mu r - r^3,$$
$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = -1$$

What kind of bifurcation takes place at $\mu = 0$? Where is the limit cycle in this system, and is it stable or unstable? Draw the phase plane diagrams for the cases $\mu < 0$ and $\mu > 0$.

3

A bit of group theory

This book is about patterns that have some kind of symmetry The natural way to describe symmetry is using group theory, so we had better know something about groups before we start to investigate patterns This chapter gives you a quick introduction to the concept of a group and some relevant bits of algebra. If you would like a bit more background reading you might like to try Cornwell (1984) for an introduction to group theory in physics, James and Liebeck (1993) for more depth on representations and characters, Steeb (1996) for Lie groups and continuous symmetries, and Johnston and Richman (1997) or Cohn (1982) for a general introductory algebra text.

3.1 Groups

A **group**, Γ, consists of a set of elements $\{\gamma_1, \gamma_2, \ldots\}$, together with an operation (known generically as multiplication or composition), that satisfy the group axioms below

Composition associates with every pair of elements γ_1 and γ_2 of Γ another element γ_3 of Γ according to

$$\gamma_3 = \gamma_1\gamma_2, \tag{3.1}$$

so any group must be **closed** under the group operation. Sometimes the symbol $\circ$ or $\times$ is used for composition, for example $\gamma_1 \circ \gamma_2$ or $\gamma_1 \times \gamma_2$ Composition is 'on the left': in other words $\gamma_1\gamma_2$ means 'do γ_2 first, followed by γ_1'

The group axioms are:

(i) *Closure*
The group must be closed under composition.

(ii) *Associativity*
Any three elements γ_1, γ_2 and γ_3 of Γ must satisfy

$$(\gamma_1\gamma_2)\gamma_3 = \gamma_1(\gamma_2\gamma_3) \tag{3.2}$$

(iii) *Identity*
There must be an **identity** element, $e \in \Gamma$, such that

$$\gamma e = e\gamma = \gamma, \quad \forall \gamma \in \Gamma \tag{3.3}$$

(iv) *Inverses*
For each element $\gamma \in \Gamma$ there must be an inverse element, $\gamma^{-1} \in \Gamma$, such that

$$\gamma\gamma^{-1} = \gamma^{-1}\gamma = e \tag{3.4}$$

Example 3.1 (The multiplicative group of nonzero real numbers) *Consider the nonzero real numbers under multiplication. Checking the group axioms in turn we see that*

(i) *The group is closed since real numbers r_i multiply together according to $r_3 = r_1 r_2$, and r_3 will be nonzero and real if r_1 and r_2 are.*
(ii) *Ordinary multiplication is associative, since $(r_1 r_2)r_3 = r_1(r_2 r_3) = r_1 r_2 r_3$*
(iii) *The identity element is simply 1, since $1 \times r = r \times 1 = r$ for any real number r, and 1 is a nonzero real number.*
(vi) *The inverse of a nonzero real number r is $1/r$ (also a nonzero real number) since $r \times (1/r) = (1/r) \times r = 1$*

The group axioms hold and so the nonzero real numbers under multiplication form a group.

Example 3.2 (The additive group of real numbers) *This time we take the group operation to be addition. Clearly the group is closed, and addition is associative. The identity element is 0, since $r + 0 = 0 + r = r$ for any real number, r, and the inverse of a number r is $-r$, since $r + (-r) = -r + r = 0$. Again the group axioms hold, and thus the real numbers also form a group under addition.*

Example 3.3 (The general linear group, $GL(n)$) *The set of all real invertible $n \times n$ matrices forms the general linear group of degree n, $GL(n)$, under matrix multiplication. The group is closed since AB is a real $n \times n$ matrix if A and B are. Matrix multiplication is associative, there is an identity matrix I_n, and each matrix has an inverse, so $GL(n)$ satisfies the group axioms.*

Strictly speaking we should write $GL(n, \mathbb{R})$ for the group we have just defined, and call it the general linear group of degree n over $\mathbb{R}$, since we can also define the general linear group $GL(n, F)$ of degree n over any other field F. In particular you may come across $GL(n, \mathbb{C})$, which is the group of invertible complex $n \times n$ matrices under matrix multiplication. However, we shall use $GL(n)$ as shorthand for $GL(n, \mathbb{R})$ throughout this book.

An **Abelian** group is one where all the group elements commute:

$$\gamma_1\gamma_2 = \gamma_2\gamma_1, \quad \forall\gamma_1, \gamma_2 \in \Gamma. \tag{3.5}$$

For example, the group of real numbers under addition is abelian, since we have

$$a + b = b + a, \quad \forall a, b \in \mathbb{R} \tag{3.6}$$

The group operation of Abelian groups is sometimes written as addition (+) to stress the commutativity, even though it may actually be multiplication or composition of symmetries. We won't use this notation in this book, but you might come across it in research papers.

Many groups are not Abelian, and for these groups the order of composition of the elements is important. For example, $GL(n)$ is not Abelian, because matrix multiplication does not commute, as the following calculation shows:

$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.7}$$

Most of the groups we shall be concerned with in this book are not Abelian. Since the order of composition of group elements matters, it is important to remember that the inverse of a group element $\gamma_1\gamma_2$ is $\gamma_2^{-1}\gamma_1^{-1}$, because

$$\gamma_1\gamma_2\gamma_2^{-1}\gamma_1^{-1} = \gamma_1 e\gamma_1^{-1} = e = \gamma_2^{-1}e\gamma_2 = \gamma_2^{-1}\gamma_1^{-1}\gamma_1\gamma_2, \ \forall\gamma_1, \gamma_2 \in \Gamma. \tag{3.8}$$

Products of a group element with itself are written as powers of that element. For example:

$$\gamma\ \gamma = \gamma^2, \tag{3.9}$$

$$\gamma^2\ \gamma^3 = \gamma^5, \tag{3.10}$$

$$\gamma^{-1}\ \gamma^{-1} = \gamma^{-2}, \tag{3.11}$$

$$\gamma^0 = e. \tag{3.12}$$

A group is said to be **generated** by a subset of its elements, the **generators**, if all group elements can be written as products of powers of the generators.

Example 3.4 (The cyclic group, C_n, of order n) *Consider the group generated by $a \equiv e^{2\pi i/n} \in \mathbb{C}$ under multiplication. The group elements are $\{e, a, a^2, \ldots, a^{n-1}\}$, where the identity element is given by $e = 1$, and $a^n = e$. The group is closed since $a^p \in \{e, a, a^2, \ldots, a^{n-1}\}$ for any integer, p, and there are inverses $(a^p)^{-1} = a^{n-p}$, for $p = 0, \ldots, n-1$. Finally multiplication is associative, so all the group axioms hold. This is C_n, the cyclic group of order n.*

The generator a does not have to be $e^{2\pi i/n}$: any group with elements $\{e, a, a^2, \ldots, a^{n-1}\}$, where $a^n = e$, is cyclic of order n. In fact, strictly speaking,

Table 3.1 *A group table*

	e	γ_1	γ_2	...	γ_n
e	e	γ_1	γ_2	...	γ_n
γ_1	γ_1	γ_1^2	$\gamma_1\gamma_2$	...	$\gamma_1\gamma_n$
γ_2	γ_2	$\gamma_2\gamma_1$	γ_2^2	...	$\gamma_2\gamma_n$
...	...	...	...	...	...
γ_n	γ_n	$\gamma_n\gamma_1$	$\gamma_n\gamma_2$	...	γ_n^2

C_n should be defined in this abstract way. We will often come across these groups when a is the rotation through $2\pi/n$.

The **order**, n, of a group element $\gamma \in \Gamma$ is the smallest integer such that $\gamma^n = e$. For example, the order of the generator a of the cyclic group C_4 is 4, and the order of $a^2 \in C_4$ is 2.

If a group has a finite number of elements it is called a **finite group**. The **order**, $|\Gamma|$, of a finite group Γ is the number of elements of Γ. For example, the additive and multiplicative groups of real numbers are both infinite groups, while the cyclic group, C_n, is finite, with n elements, so its order is n.

The composition of group elements of a finite group is sometimes displayed in a **group table**, like Table 3.1. Each entry in the body of a group table is the composition, $\gamma_1\gamma_2$ of the element γ_1 in the corresponding row of the lefthand column of the table with the element γ_2 in the corresponding column of the top row.

Example 3.5 (The symmetry group of an equilateral triangle) *$\mathcal{D}_3$, the symmetry group of an equilateral triangle is generated by ρ, a rotation through an angle of $2\pi/3$, and m, a reflection in a line joining a vertex to the midpoint of the opposite side.*

The group elements are $e, \rho, \rho^2, m, m\rho, m\rho^2$, corresponding to the identity, rotations through $2\pi/3$ and $4\pi/3$, and three reflections respectively. The symmetries are shown in Figure 3.1. The group table (Table 3.2) shows that performing any two of these symmetries one after the other gives another symmetry in the group, so the group is closed. For example $\rho \circ m\rho = m$, where $\circ$ represents composition of symmetries. Recall that in composition the symmetry on the righthand side is performed before the symmetry on the lefthand side. You can also see from the group table that $\mathcal{D}_3$ is not Abelian, for example $m \circ \rho \neq \rho \circ m = m\rho^2$.

Composition of symmetries is associative. The identity element is e (the symmetry of doing nothing!), and all symmetries have inverses that are also group elements. For example $\rho \circ \rho^2 = \rho^2 \circ \rho = e$, so $\rho^{-1} = \rho^2$.

Table 3.2 *The group table for $\mathcal{D}_3$, the symmetry group of an equilateral triangle*

	e	ρ	ρ^2	m	$m\rho$	$m\rho^2$
e	e	ρ	ρ^2	m	$m\rho$	$m\rho^2$
ρ	ρ	ρ^2	e	$m\rho^2$	m	$m\rho$
ρ^2	ρ^2	e	ρ	$m\rho$	$m\rho^2$	m
m	m	$m\rho$	$m\rho^2$	e	ρ	ρ^2
$m\rho$	$m\rho$	$m\rho^2$	m	ρ^2	e	ρ
$m\rho^2$	$m\rho^2$	m	$m\rho$	ρ	ρ^2	e

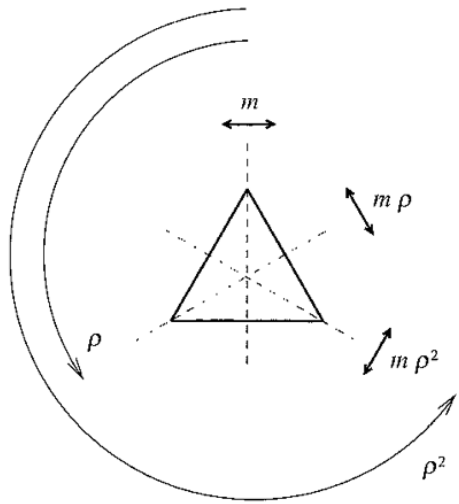


Fig. 3.1. The elements of $\mathcal{D}_3$

Example 3.6 (The dihedral group, $\mathcal{D}_n$, of order $2n$) *The symmetry group of a regular polygon with n sides is known as $\mathcal{D}_n$, the dihedral group of order $2n$, and can be treated in a similar way to the symmetry group of the equilateral triangle in the previous example. The group is generated by a rotation, ρ, through $2\pi/n$, and a reflection, m, with an axis that passes through the centre of the polygon and a corner or the midpoint of a side. The $2n$ group elements are the identity, the rotations $\{\rho, \ldots, \rho^{n-1}\}$ and the reflections $\{m, m\rho, \ldots, m\rho^{n-1}\}$.*

Example 3.7 (The Euclidean group, $E(2)$) *The Euclidean group, $E(2)$, is the group of translations, reflections and rotations of the plane. For a transformation $(\rho, \boldsymbol{t})$, consisting of ρ, a reflection in an axis containing the origin, or a rotation*

about the origin, followed by $\boldsymbol{t}$*, a translation, the vector* $\boldsymbol{x} \in \mathbb{R}^2$ *is transformed to*

$$\boldsymbol{x}' = (\rho, \boldsymbol{t})\boldsymbol{x} \equiv \rho\boldsymbol{x} + \boldsymbol{t} \tag{3.13}$$

Performing transformation $(\rho_1, \boldsymbol{t}_1)$ *after transformation* $(\rho_2, \boldsymbol{t}_2)$ *gives*

$$\boldsymbol{x}'' = \rho_1\rho_2\boldsymbol{x} + \rho_1\boldsymbol{t}_2 + \boldsymbol{t}_1, \tag{3.14}$$

which is equivalent to a transformation $(\rho, \boldsymbol{t})$ *with* $\rho = \rho_1\rho_2$ *and* $\boldsymbol{t} = \rho_1\boldsymbol{t}_2 + \boldsymbol{t}_1$*, so the group is closed. Composition of transformations is also associative*

$$\begin{aligned}(\rho_1, \boldsymbol{t}_1)((\rho_2, \boldsymbol{t}_2)(\rho_3, \boldsymbol{t}_3)) &= (\rho_1, \boldsymbol{t}_1)(\rho_2\rho_3, \rho_2\boldsymbol{t}_3 + \boldsymbol{t}_2) \\ &= (\rho_1\rho_2\rho_3, \rho_1(\rho_2\boldsymbol{t}_3 + \boldsymbol{t}_2) + \boldsymbol{t}_1) \\ &= (\rho_1\rho_2\rho_3, \rho_1\rho_2\boldsymbol{t}_3 + \rho_1\boldsymbol{t}_2 + \boldsymbol{t}_1) \\ &= (\rho_1\rho_2, \rho_1\boldsymbol{t}_2 + \boldsymbol{t}_1)(\rho_3, \boldsymbol{t}_3) \\ &= ((\rho_1, \boldsymbol{t}_1)(\rho_2, \boldsymbol{t}_2))(\rho_3, \boldsymbol{t}_3).\end{aligned} \tag{3.15}$$

There is an identity element $(e, \mathbf{0})$*, and an element* $(\rho, \boldsymbol{t})$ *has inverse* $(\rho^{-1}, -\rho^{-1}\boldsymbol{t})$ *All the group axioms are satisfied and so* $E(2)$ *is a group.*

The dihedral groups and the Euclidean group are important in pattern formation, as we shall see in later chapters

3.2 Subgroups, quotient groups and conjugacy

A **subgroup**, H, of a group, Γ, is a subset of the group elements that forms a group under the same group operation as Γ

To show that a subset of group elements, $H \subseteq \Gamma$, is a subgroup you need only check that it is closed under composition and contains the identity element and the inverses of all its elements. Associativity under composition must hold, since Γ is a group

H is a **normal** subgroup of a group, Γ, if

$$\gamma h \gamma^{-1} \in H, \quad \forall \gamma \in \Gamma, \forall\, h \in H. \tag{3.16}$$

So for a normal subgroup we have $\gamma H \gamma^{-1} \subseteq H$, $\forall \gamma \in \Gamma$ However we also have

$$\gamma^{-1} h \gamma \in H, \quad \forall \gamma \in \Gamma, \ \forall h \in H, \tag{3.17}$$

and so $h \in \gamma H \gamma^{-1}$. In other words we have $H \subseteq \gamma H \gamma^{-1}$, $\forall \gamma \in \Gamma$. Since the inclusion holds both ways round we must have

$$\gamma H \gamma^{-1} = H \tag{3.18}$$

If Γ is Abelian then all subgroups are normal since $\gamma h \gamma^{-1} = \gamma\gamma^{-1}h = h$

The **complement**, $\Gamma \backslash H$ or $\Gamma - H$, of a subgroup H in Γ is the set of all group elements that do not belong to H:

$$\Gamma - H = \{\gamma \in \Gamma : \gamma \neq H\} \tag{3.19}$$

The complement of a subgroup is not a subgroup since it cannot contain the identity element

Example 3.8 *C_3, the group of rotations of an equilateral triangle, consisting of the elements $\{e, \rho, \rho^2\}$ under composition, is a subgroup of $\mathcal{D}_3$. Looking at the group table, Table 3.2, you can see that these elements when composed with each other always give another element in the subgroup. C_3 contains the identity e, and all the inverses are contained in the subgroup since $\rho^{-1} = \rho^2$ and $\rho^{-2} = \rho$. The complement $\mathcal{D}_3 - C_3$ consists of the reflections $\{m, m\rho, m\rho^2\}$*

More generally, the cyclic group C_n of rotations generated by rotation through $2\pi/n$ is always a subgroup of the dihedral group $\mathcal{D}_n$.

Example 3.9 *It is straightforward to check that the pure translations, $(e, \boldsymbol{t})$, form a subgroup of the Euclidean group $E(2)$, as do the reflections and rotations, $(\rho, \mathbf{0})$. The point of this example is to see whether these are normal subgroups of $E(2)$.*

For any two group elements, $(\rho_1, \boldsymbol{t}_1)$ and $(\rho_2, \boldsymbol{t}_2)$, in $E(2)$ we have

$$\begin{aligned}(\rho_1, \boldsymbol{t}_1)(\rho_2, \boldsymbol{t}_2)(\rho_1, \boldsymbol{t}_1)^{-1} &= (\rho_1, \boldsymbol{t}_1)(\rho_2, \boldsymbol{t}_2)(\rho_1^{-1}, -\rho_1^{-1}\boldsymbol{t}_1) \\ &= (\rho_1\rho_2, \rho_1\boldsymbol{t}_2 + \boldsymbol{t}_1)(\rho_1^{-1}, -\rho_1^{-1}\boldsymbol{t}_1) \\ &= (\rho_1\rho_2\rho_1^{-1}, -\rho_1\rho_2\rho_1^{-1}\boldsymbol{t}_1 + \rho_1\boldsymbol{t}_2 + \boldsymbol{t}_1) \end{aligned} \tag{3.20}$$

If $(\rho_2, \boldsymbol{t}_2)$ is a pure translation, then $\rho_2 = e$ and we have

$$\begin{aligned}(\rho_1, \boldsymbol{t}_1)(\rho_2, \boldsymbol{t}_2)(\rho_1, \boldsymbol{t}_1)^{-1} &= (\rho_1, \boldsymbol{t}_1)(e, \boldsymbol{t}_2)(\rho_1, \boldsymbol{t}_1)^{-1} \\ &= (\rho_1 e\rho_1^{-1}, -\rho_1 e\rho_1^{-1}\boldsymbol{t}_1 + \rho_1\boldsymbol{t}_2 + \boldsymbol{t}_1) \\ &= (e, -\boldsymbol{t}_1 + \rho_1\boldsymbol{t}_2 + \boldsymbol{t}_1) \\ &= (e, \rho_1\boldsymbol{t}_2), \end{aligned} \tag{3.21}$$

so $(\rho_1, \boldsymbol{t}_1)(e, \boldsymbol{t}_2)(\rho_1, \boldsymbol{t}_1)^{-1}$ is also a pure translation for all $(\rho_1, \boldsymbol{t}_1) \in E(2)$. This means the group of translations is a normal subgroup of $E(2)$. However, if $(\rho_2, \boldsymbol{t}_2)$ is a pure reflection or rotation, then $\boldsymbol{t}_2 = \mathbf{0}$, and we have

$$\begin{aligned}(\rho_1, \boldsymbol{t}_1)(\rho_2, \boldsymbol{t}_2)(\rho_1, \boldsymbol{t}_1)^{-1} &= (\rho_1, \boldsymbol{t}_1)(\rho_2, \mathbf{0})(\rho_1, \boldsymbol{t}_1)^{-1} \\ &= (\rho_1\rho_2\rho_1^{-1}, -\rho_1\rho_2\rho_1^{-1}\boldsymbol{t}_1 + \boldsymbol{t}_1). \end{aligned} \tag{3.22}$$

In general $\rho_1\rho_2\rho_1^{-1}\boldsymbol{t}_1 \neq \boldsymbol{t}_1$ *and so* $(\rho_1, \boldsymbol{t}_1)(\rho_2, \boldsymbol{0})(\rho_1, \boldsymbol{t}_1)^{-1}$ *is not a pure reflection or rotation, and the group of reflections and rotations is not a normal subgroup of* $E(2)$.

The **normalizer**, $N(H)$, of a subgroup H of Γ is given by

$$N(H) = \{\gamma \in \Gamma : \gamma^{-1}H\gamma = H\} \tag{3 23}$$

Clearly $H \subseteq N(H)$, and H is normal in $N(H)$. In fact the normalizer is the largest subgroup of Γ that has H as a normal subgroup. To show that $N(H)$ is a subgroup, we check for closure, the identity and inverses. $N(H)$ is closed under composition, since

$$(\gamma_1\gamma_2)^{-1}H\gamma_1\gamma_2 = \gamma_2^{-1}\gamma_1^{-1}H\gamma_1\gamma_2 = \gamma_2^{-1}H\gamma_2 = H, \quad \forall\gamma_1, \gamma_2 \in \Gamma, \tag{3.24}$$

and so $\gamma_1\gamma_2 \in N(H)$ if γ_1 and γ_2 are in $N(H)$. We also have an identity element, since $e \in N(H)$, and inverses since $\gamma^{-1} \in N(H)$ if $\gamma \in N(H)$ because $\gamma^{-1}H\gamma = H$ implies $H = \gamma H\gamma^{-1}$. So $N(H)$ is indeed a subgroup. It must be the largest subgroup within which H is normal, since any element $\gamma \in \Gamma - N(H)$ must satisfy $\gamma^{-1}H\gamma \neq H$, and so H cannot be normal within a subgroup that includes any elements outside $N(H)$

Example 3.10 *The reflection subgroup* $\{e, m\}$ *of* $\mathcal{D}_3$ *is its own normalizer, while the normalizer of the subgroup of rotations* $\mathcal{C}_3 = \{e, \rho, \rho^2\} \subset \mathcal{D}_3$ *is the whole of* $\mathcal{D}_3$.

If H is a subgroup of a group Γ, then the **left coset** containing $\gamma \in \Gamma$ is defined by

$$\gamma H = \{\gamma h | h \in H\}. \tag{3 25}$$

The **right coset** is defined similarly to be

$$H\gamma = \{h\gamma | h \in H\} \tag{3 26}$$

(Note that some authors call the left coset, as defined here, the right coset and vice versa. As long as you stick to one convention, it doesn't matter.) For a normal subgroup we have $\gamma H\gamma^{-1} = H$ or

$$\gamma H = H\gamma, \quad \forall\gamma \in \Gamma, \tag{3 27}$$

which says that the left and right cosets are the same if H is a normal subgroup. In this case we can call them simply cosets.

Cosets are disjoint Consider the left cosets $\gamma_1 H$ and $\gamma_2 H$ If they have an element in common then there exist $h_1, h_2 \in H$ such that

$$\gamma_1 h_1 = \gamma_2 h_2, \tag{3.28}$$

and so

$$\gamma_1^{-1}\gamma_2 = h_1 h_2^{-1} \in H \tag{3.29}$$

Thus

$$\gamma_2 h = \gamma_1(\gamma_1^{-1}\gamma_2)h \in \gamma_1 H, \quad \forall h \in H, \tag{3.30}$$

and we have $\gamma_2 H \subseteq \gamma_1 H$. But using an identical argument we also have $\gamma_1 H \subseteq \gamma_2 H$, and so $\gamma_1 H = \gamma_2 H$ So the left cosets are disjoint and we have

$$\gamma_1 H = \gamma_2 H \quad \text{or} \quad \gamma_1 H \cap \gamma_2 H = \emptyset. \tag{3.31}$$

Similarly the right cosets are disjoint. The cosets therefore partition the group Γ.

If H is a normal subgroup of a group, Γ, the **quotient group**, Γ/H, has group elements given by the set of distinct cosets, and a group operation defined by

$$(\gamma_1 H)(\gamma_2 H) = \gamma_1\gamma_2 H, \quad \forall \gamma_1, \gamma_2 \in \Gamma. \tag{3.32}$$

Composition is well-defined since

$$\begin{aligned}(\gamma_1 H)(\gamma_2 H) &= \gamma_1(\gamma_2\gamma_2^{-1})H\gamma_2 H, \quad (\text{since } \gamma_2\gamma_2^{-1} = e) && (3.33)\\ &= \gamma_1\gamma_2(\gamma_2^{-1}H\gamma_2)H, && (3.34)\\ &= \gamma_1\gamma_2 H H, \quad (\text{since } \gamma_2^{-1}H\gamma_2 = H) && (3.35)\\ &= \gamma_1\gamma_2 H && (3.36)\end{aligned}$$

Γ/H is closed under composition and associative because Γ is The identity is $eH = H$ and the inverse of γH is $\gamma^{-1}H$.

Example 3.11 ($\mathbb{Z}/n$, **the integers under addition modulo** n) *Let Γ be the group of integers, $\mathbb{Z}$, under addition, and let H be $n\mathbb{Z}$, the subgroup of integers divisible by n. $n\mathbb{Z}$ is normal in $\mathbb{Z}$, since the group is Abelian. The distinct cosets are $\{i + H,\ 0 \le i \le n-1\}$, and the group operation on the quotient group $\mathbb{Z}/n\mathbb{Z}$ is*

$$i + H + j + H = (i + j \mod n) + H. \tag{3.37}$$

The group $\mathbb{Z}/n\mathbb{Z}$ is often written $\mathbb{Z}/n$ or $\mathbb{Z}_n$ It is a cyclic group of order n, generated by the element 1, giving rise to the common alternative notation $\mathbb{Z}_n$ for C_n. The notation $\mathbb{Z}_2$ is very frequently used for the cyclic group of order 2

The quotient group is useful in this example because it 'factors out' the repetitive behaviour of the larger group, leaving only the essential structure If we are

looking at addition modulo 10 then the numbers n, 10 + n, 20 + n and so on, for integers n between 0 and 9, all behave in the same way, so we might as well restrict our attention to the integers 0, 1, . . . , 9 and forget about the rest. The quotient group does this for us. Quotient groups are similarly useful in describing periodic patterns, by allowing us to focus on the basic unit of the pattern rather than having to worry about all the repeats

Two group elements, h_1 and h_2, in Γ are **conjugate**, or equivalently are said to be in the same **class** or **conjugacy class** if there is a group element, $\gamma \in \Gamma$, such that

$$h_1 = \gamma h_2 \gamma^{-1} \tag{3.38}$$

Any two group elements in the same conjugacy class have the same order, since if h_2 is of order n, and $h_1 = \gamma h_2 \gamma^{-1}$ for some $\gamma \in \Gamma$ then

$$h_1^n = (\gamma h_2 \gamma^{-1})^n \tag{3.39}$$

$$= \gamma h_2 \gamma^{-1} \gamma h_2 \gamma^{-1} \ldots \gamma h_2 \gamma^{-1} \tag{3.40}$$

$$= \gamma h_2^n \gamma^{-1} = \gamma e \gamma^{-1} = e, \tag{3.41}$$

using $h_2^n = e$, and so h_1 is also of order n.

Two subgroups H_1 and H_2 of Γ are **conjugate** if there is a group element $\gamma \in \Gamma$ such that

$$H_1 = \gamma H_2 \gamma^{-1}. \tag{3.42}$$

Example 3.12 *In D_3, the elements ρ and ρ^2 are conjugate since $\rho^2 = m\rho m^{-1}$. Both ρ and ρ^2 have order three. Similarly the reflection elements m, $m\rho$ and $m\rho^2$ are all conjugate since $\rho m \rho^{-1} = m\rho$ and $\rho^2 m(\rho^2)^{-1} = m\rho^2$. The reflections are all order two.*

The subgroups $\{e, m\}$ and $\{e, m\rho\}$ are conjugate since $\rho m \rho^{-1} = m\rho$ and $\rho e \rho^{-1} = e$.

The symmetry groups of patterns that are similar, in a sense to be made precise in Chapter 4, turn out to be conjugate.

3.3 Mappings of groups

It is possible to map one group to another. This will be useful later for characterising the action of the group on the space of pattern solutions.

A **homomorphism** is a map $\theta : \Gamma \to \Delta$ between groups Γ and Δ such that

$$\theta(\gamma_1 \gamma_2) = \theta(\gamma_1)\theta(\gamma_2), \quad \forall\, \gamma_1, \gamma_2 \in \Gamma \tag{3.43}$$

In other words homomorphisms are maps that respect the group structure A homomorphism always maps the identity element $e_\Gamma \in \Gamma$ to the identity element $e_\Delta \in \Delta$, since

$$\theta(e_\Gamma)\theta(\gamma) = \theta(e_\Gamma \gamma) = \theta(\gamma), \quad \forall \gamma \in \Gamma. \tag{3.44}$$

Multiplying on the righthand side by $(\theta(\gamma))^{-1}$ (since the group element $\theta(\gamma) \in \Delta$ must have an inverse) we find

$$\theta(e_\Gamma) = e_\Delta \tag{3.45}$$

Inverses are also mapped in the obvious way since

$$\theta(\gamma)\theta(\gamma^{-1}) = \theta(\gamma^{-1})\theta(\gamma) = \theta(e_\Gamma) = e_\Delta, \quad \forall \gamma \in \Gamma, \tag{3.46}$$

and so $(\theta(\gamma))^{-1} = \theta(\gamma^{-1})$.

Example 3.13 *Let Γ be the positive real numbers under multiplication, and let Γ' be $\mathbb{1}$, the group consisting solely of the identity element Then $\theta : \mathbb{R}^+ \to \mathbb{1}$ defined by $\theta(x) = e$, $\forall x \in \mathbb{R}^+$ is a homomorphism since*

$$\theta(xy) = e = e^2 = \theta(x)\theta(y), \quad \forall x, y \in \mathbb{R}^+. \tag{3.47}$$

This point of this rather boring example is to show that homomorphisms need not be one-to-one. We will come to some more interesting homomorphisms in a minute.

An **isomorphism**, θ, between groups Γ and Δ is a homomorphism that is bijective. If such an isomorphism exists, we then say that Γ and Δ are isomorphic, written $\Gamma \cong \Delta$.

Example 3.14 *The cyclic group of order n generated by $e^{2\pi i/n}$ under multiplication, with group elements $\{1, e^{2\pi i/n}, \ldots, e^{2\pi i(n-1)/n}\}$ is isomorphic to the symmetry group generated by a rotation, ρ, through $2\pi/n$ under composition of symmetries, with group elements $\{e, \rho, \ldots, \rho^{n-1}\}$ The map, θ, defined by*

$$\theta(e^{2\pi i m/n}) = \rho^m, \quad m \in \mathbb{Z},\ 0 \le m \le n-1 \tag{3.48}$$

is clearly bijective and it respects the group structure since

$$\theta(e^{2\pi i m/n} e^{2\pi i p/n}) = \rho^{(m+p)} = \rho^m \rho^p = \theta(e^{2\pi i m/n})\theta(e^{2\pi i p/n}),$$
$$\forall\, m, p \in \mathbb{Z} :\ 0 \le m, p \le n-1 \tag{3.49}$$

So θ is an isomorphism, and the two groups are isomorphic as claimed.

Example 3.15 (The orthogonal groups, $O(n)$, and special orthogonal groups, $SO(n)$) *The n-dimensional orthogonal group, $O(n)$, consists of all $n \times n$*

matrices, M, that satisfy

$$MM^T = I_n, \tag{3 50}$$

where M^T is the transpose of M $O(n)$ is isomorphic to the group of all rotations and reflections in $\mathbb{R}^n$ that fix the origin, with a mapping between each matrix and the transformation that it represents in coordinate geometry Similarly, the n-dimensional special orthogonal group, $SO(n)$, consists of matrices $M \in O(n)$ that satisfy $\det M = 1$ *$SO(n)$ is isomorphic to the group of all rotations in $\mathbb{R}^n$ that fix the origin, and is often called the n-dimensional rotation group The group of rotations in the plane is often called the* **circle group**, *S^1, and is isomorphic to $SO(2)$.*

Isomorphic groups have the same abstract form, so their names are sometimes used interchangeably, for example we might say that $O(2)$ is the group of reflections and rotations of the plane that keep the origin fixed, and $SO(2)$ is the group of rotations of the plane that keep the origin fixed When it comes to describing how isomorphic groups act on a space, though, we may have to distinguish between them. For example, the group generated by a reflection and the group generated by a rotation through π are both isomorphic to the cyclic group $\mathbb{Z}_2$, but they act on the plane in different ways We will go into this in more depth when we consider representations of groups later on

An isomorphism from a group Γ to itself is called an **automorphism** of Γ.

Example 3.16 *Let Γ be the real numbers under addition. Then $\theta : \mathbb{R} \to \mathbb{R}$ defined by $\theta(x) = 2x$ is an automorphism. The map, θ, is a homomorphism since we have*

$$\theta(x + y) = 2(x + y) = 2x + 2y = \theta(x)\theta(y), \quad \forall\, x, y \in \mathbb{R} \tag{3.51}$$

It is an automorphism because it is bijective (the inverse exists and is given by $\theta^{-1}(x) = x/2$, $\forall x \in \mathbb{R}$), and maps the group of real numbers under addition onto itself.

Aut(Γ) is the set of all automorphisms of Γ and forms a group under composition of functions The group is closed since the composition of two automorphisms is also an automorphism Composition of functions is associative There is an identity automorphism, $Id : \Gamma \to \Gamma$, defined by $Id(\gamma) = \gamma, \forall \gamma \in \Gamma$, and every automorphism has an inverse, since automorphisms are bijective. Thus the group axioms hold and Aut(Γ) under composition of functions does form a group

3.4 Products of groups

The **direct product**, $\Gamma \times \Delta$, of two groups Γ and Δ has group elements (γ, δ), where $\gamma \in \Gamma$ and $\delta \in \Delta$. The group operation is defined by

$$(\gamma_1, \delta_1)(\gamma_2, \delta_2) = (\gamma_1\gamma_2, \delta_1\delta_2). \tag{3.52}$$

$\Gamma \times \Delta$ is closed, since Γ and Δ are. The group operation on $\Gamma \times \Delta$ is associative since the group operations on Γ and Δ are. The identity element is (e_Γ, e_Δ), where e_Γ is the identity element in Γ and e_Δ is the identity element in Δ, and an element (γ, δ) has inverse $(\gamma^{-1}, \delta^{-1})$. So all the group axioms are satisfied and $\Gamma \times \Delta$ is a group.

The direct product of Abelian groups is often written as a **direct sum**, $\Gamma \oplus \Delta$. The use of the plus sign here mirrors its use to represent the group operation of an Abelian group. Again we will not use this notation, but you may see it in research papers.

The group $\Gamma \times \Delta$ has normal subgroups (Γ, e_Δ) and (e_Γ, Δ) isomorphic to Γ and Δ, respectively, since

$$\begin{aligned}(\gamma, \delta)(\Gamma, e_\Delta)(\gamma, \delta)^{-1} &= (\gamma\Gamma\gamma^{-1}, \delta e_\Delta \delta^{-1}) \\ &= (\Gamma, e_\Delta), \quad \forall\, (\gamma, \delta) \in \Gamma \times \Delta, \qquad (3.53)\\ (\gamma, \delta)(e_\Gamma, \Delta)(\gamma, \delta)^{-1} &= (\gamma e_\Gamma \gamma^{-1}, \delta\Delta\delta^{-1}) \\ &= (e_\Gamma, \Delta), \quad \forall\, (\gamma, \delta) \in \Gamma \times \Delta. \qquad (3.54)\end{aligned}$$

The direct product generalises to products of more than two groups in the natural way. For example, we would define $\Gamma_1 \times \Gamma_2 \times \Gamma_3$ by

$$\begin{aligned}&(\gamma_1, \gamma_2, \gamma_3)(\gamma_1', \gamma_2', \gamma_3') = (\gamma_1\gamma_1', \gamma_2\gamma_2', \gamma_3\gamma_3'), \\ &\forall \gamma_1, \gamma_1' \in \Gamma_1,\ \forall \gamma_2, \gamma_2' \in \Gamma_2,\ \forall \gamma_3, \gamma_3' \in \Gamma_3 \qquad (3.55)\end{aligned}$$

Example 3.17 *The symmetry group of a rectangle, $\mathcal{D}_2$, is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$. $\mathcal{D}_2$ is generated by two commuting reflections, m_x and m_y, shown in Figure 3.2. The group elements are $\{e, m_x, m_y, m_x m_y\}$, where $m_x m_y (= m_y m_x)$ is the rotation through an angle π. Since the reflections commute, the group is Abelian. The reflection group $\mathbb{Z}_2$ has elements $\{e, m\}$, where m is any reflection, and $m^2 = e$. If we have two copies of $\mathbb{Z}_2$, one generated by m_x and one by m_y, then the direct product group $\mathbb{Z}_2 \times \mathbb{Z}_2$ has elements $\{(e, e), (m_x, e), (e, m_y), (m_x, m_y)\}$. We can define the isomorphism $\theta : \mathcal{D}_2 \to \mathbb{Z}_2 \times \mathbb{Z}_2$ by*

$$\theta(e) = (e, e), \tag{3.56}$$

$$\theta(m_x) = (m_x, e), \tag{3.57}$$

$$\theta(m_y) = (e, m_y), \tag{3.58}$$

$$\theta(m_x m_y) = (m_x, m_y), \tag{3.59}$$

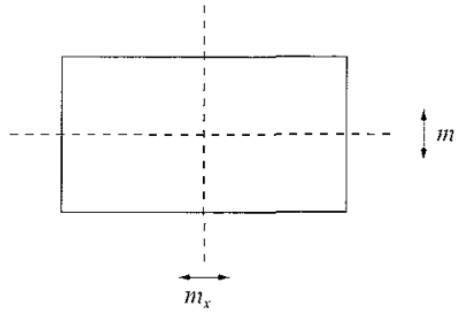


Fig 3.2 The generators of D_2

so we have $D_2 \cong \mathbb{Z}_2 \times \mathbb{Z}_2$ *The group* $\mathbb{Z}_2 \times \mathbb{Z}_2$ *is often denoted* $\mathbb{Z}_2^2$ *Similarly we write* $\mathbb{Z}_2^n = \mathbb{Z}_2 \times \mathbb{Z}_2 \times \quad . \times \mathbb{Z}_2$ *for a direct product of n copies of* $\mathbb{Z}_2$

The **semidirect product**, $\Gamma \ltimes \Delta$, of two groups Γ and Δ has group elements (γ, δ), where $\gamma \in \Gamma$ and $\delta \in \Delta$. If $\theta : \Gamma \to \mathrm{Aut}(\Delta)$ is a group homomorphism, then the semidirect product of Γ and Δ with respect to θ is defined by the group operation

$$(\gamma_1, \delta_1)(\gamma_2, \delta_2) = (\gamma_1\gamma_2, \delta_1\theta(\gamma_1)(\delta_2)) \tag{3.60}$$

The semidirect product is not unique as it depends on the choice of homomorphism θ It is often written as a **semidirect sum**, $\Gamma \dot{+} \Delta$, when Δ is Abelian and the group operation on Δ can be written as addition Lattice groups, relevant to the description of periodic patterns, have a semidirect product structure.

To show that $\Gamma \ltimes \Delta$ is a group, we check that it satisfies the group axioms. $\Gamma \ltimes \Delta$ is closed since $\gamma_1\gamma_2 \in \Gamma$, for all $\gamma_1, \gamma_2 \in \Gamma$, and by the definition of θ we have $\theta(\gamma_1)(\delta_2) \in \Delta$, and so $\delta_1\theta(\gamma_1)(\delta_2) \in \Delta$, for all $\gamma_1 \in \Gamma$ and $\delta_1, \delta_2 \in \Delta$

There is an identity element, (e_Γ, e_Δ), satisfying

$$(e_\Gamma, e_\Delta)(\gamma, \delta) = (e_\Gamma\gamma, e_\Delta\theta(e_\Gamma)\delta) = (\gamma, \delta), \quad \forall\gamma \in \Gamma, \forall\delta \in \Delta, \tag{3.61}$$

since $\theta(e_\Gamma) = Id$, because θ is a homomorphism, and

$$(\gamma, \delta)(e_\Gamma, e_\Delta) = (\gamma e_\Gamma, \delta\theta(\gamma)e_\Delta) = (\gamma, \delta), \quad \forall\gamma \in \Gamma, \forall\delta \in \Delta, \tag{3.62}$$

since $\theta(\gamma)e_\Delta = e_\Delta$, because $\theta(\gamma)$ is an automorphism of Δ

An element (γ, δ) has inverse $(\gamma^{-1}, \theta(\gamma^{-1})(\delta^{-1}))$, since

$$\begin{aligned}(\gamma, \delta)(\gamma^{-1}, \theta(\gamma^{-1})(\delta^{-1})) &= (\gamma\gamma^{-1}, \delta\theta(\gamma)(\theta(\gamma^{-1})(\delta^{-1}))) \\ &= (e_\Gamma, \delta\ Id\ \delta^{-1}) = (e_\Gamma, e_\Delta),\end{aligned} \tag{3.63}$$

using $\theta(\gamma)\theta(\gamma^{-1}) = Id$, and we also have

$$\begin{aligned}(\gamma^{-1}, \theta(\gamma^{-1})(\delta^{-1}))(\gamma, \delta) &= (\gamma^{-1}\gamma, \theta(\gamma^{-1})(\delta^{-1})\theta(\gamma^{-1})(\delta) \\ &= (e_\Gamma, e_\Delta),\end{aligned} \tag{3.64}$$

using $\theta(\gamma^{-1})(\delta^{-1}) = (\theta(\gamma^{-1})(\delta))^{-1}$, since $\theta(\gamma^{-1})$ is an automorphism of Δ.

To show that the semidirect product, $\Gamma \ltimes \Delta$, is indeed a group it just remains to check associativity. We have

$$\begin{aligned}(\gamma_1, \delta_1)((\gamma_2, \delta_2)(\gamma_3, \delta_3)) &= (\gamma_1, \delta_1)(\gamma_2\gamma_3, \delta_2\theta(\gamma_2)(\delta_3))) \\ &= (\gamma_1\gamma_2\gamma_3, \delta_1\theta(\gamma_1)(\delta_2\theta(\gamma_2)(\delta_3))) \\ &= (\gamma_1\gamma_2\gamma_3, \delta_1\theta(\gamma_1)(\delta_2)\theta(\gamma_1)(\theta(\gamma_2)(\delta_3))),\end{aligned} \tag{3.65}$$

and

$$\begin{aligned}((\gamma_1, \delta_1)(\gamma_2, \delta_2))(\gamma_3, \delta_3) &= (\gamma_1\gamma_2, \delta_1\theta(\gamma_1)(\delta_2))(\gamma_3, \delta_3) \\ &= (\gamma_1\gamma_2\gamma_3, \delta_1\theta(\gamma_1)(\delta_2)\theta(\gamma_1\gamma_2)(\delta_3)) \\ &= (\gamma_1\gamma_2\gamma_3, \delta_1\theta(\gamma_1)(\delta_2)\theta(\gamma_1)(\theta(\gamma_2)(\delta_3))),\end{aligned} \tag{3.66}$$

so associativity holds and $\Gamma \ltimes \Delta$ satisfies all the group axioms.

$\Gamma \ltimes \Delta$ has a normal subgroup $(e_\Gamma, \Delta) \cong \Delta$ since for arbitrary $(\gamma, \delta) \in \Gamma \ltimes \Delta$ we have

$$\begin{aligned}(\gamma, \delta)(e_\Gamma, \Delta)(\gamma, \delta)^{-1} &= (\gamma, \delta)(e_\Gamma, \Delta)(\gamma^{-1}, \theta(\gamma^{-1})(\delta^{-1})) \\ &= (\gamma, \delta)(e_\Gamma\gamma^{-1}, \Delta\theta(e_\Gamma)(\theta(\gamma^{-1})(\delta^{-1}))) \\ &= (\gamma, \delta)(\gamma^{-1}, \Delta\theta(\gamma^{-1})(\delta^{-1})) \quad (\text{using } \theta(e_\Gamma) = Id) \\ &= (\gamma\gamma^{-1}, \delta\theta(\gamma)(\Delta\theta(\gamma^{-1})(\delta^{-1}))) \\ &= (e_\Gamma, \delta\theta(\gamma)(\Delta)) \quad (\text{using } \theta(\gamma^{-1})(\delta^{-1}) \in \Delta) \\ &= (e_\Gamma, \delta\Delta) = (e_\Gamma, \Delta).\end{aligned} \tag{3.67}$$

$\Gamma \ltimes \Delta$ also has a subgroup $(\Gamma, e_\Delta) \cong \Gamma$, but this is not normal in general since for arbitrary $(\gamma, \delta) \in \Gamma \ltimes \Delta$ we have

$$\begin{aligned}(\gamma, \delta)(\Gamma, e_\Delta)(\gamma, \delta)^{-1} &= (\gamma, \delta)(\Gamma, e_\Delta)(\gamma^{-1}, \theta(\gamma^{-1})(\delta^{-1})) \\ &= (\gamma, \delta)(\Gamma\gamma^{-1}, e_\Delta\theta(\Gamma)(\theta(\gamma^{-1})(\delta^{-1}))), \\ &= (\gamma, \delta)(\Gamma, \theta(\Gamma\gamma^{-1})(\delta^{-1})), \\ &= (\gamma, \delta)(\Gamma, \theta(\Gamma)(\delta^{-1})), \\ &= (\gamma\Gamma, \delta\theta(\gamma)(\theta(\Gamma)(\delta^{-1}))), \\ &= (\Gamma, \delta\theta(\gamma\Gamma)(\delta^{-1})), \\ &= (\Gamma, \delta\theta(\Gamma)(\delta^{-1})),\end{aligned} \tag{3.68}$$

and in general $\theta(\Gamma)(\delta^{-1}) \neq \delta^{-1}$, so $(\gamma, \delta)(\Gamma, e_\Delta)(\gamma, \delta)^{-1} \neq (\Gamma, e_\Delta)$

(Γ, e_Δ) could only be normal in $\Gamma \ltimes \Delta$, if $\theta(\gamma)(\delta^{-1}) = \delta^{-1}, \forall \gamma \in \Gamma, \forall \delta \in \Delta$, in which case we would have $\theta(\gamma) = Id, \forall \gamma \in \Gamma$, and $\Gamma \ltimes \Delta$ would actually be the direct product $\Gamma \times \Delta$.

Example 3.18 *The Euclidean group $E(2)$ is a semidirect product of the group $O(2)$ of rotations and reflections in the plane that fix the origin and the group of translations in the plane ($\cong \mathbb{R}^2$ under addition). A general transformation in $E(2)$ is given by $(\rho, \boldsymbol{t})$, where $\rho \in O(2)$ is a rotation or reflection and $\boldsymbol{t} \in \mathbb{R}^2$ is a translation. We saw in Example 3.7 that the product of the transformations $(\rho_1, \boldsymbol{t}_1)$ and $(\rho_2, \boldsymbol{t}_2)$ is given by*

$$(\rho_1, \boldsymbol{t}_1)(\rho_2, \boldsymbol{t}_2) = (\rho_1\rho_2, \rho_1\boldsymbol{t}_2 + \boldsymbol{t}_1). \tag{3.69}$$

Equation (3.69) shows that $E(2)$ has a semidirect product structure with $E(2) = O(2) \ltimes \mathbb{R}^2$. The homomorphism $\theta : O(2) \to \mathrm{Aut}(\mathbb{R}^2)$ is defined by $\theta(\rho)(\boldsymbol{t}) = \rho\boldsymbol{t}$, and the group operation on $\mathbb{R}^2$ is addition.

We saw in Example 3.9 that $\mathbb{R}^2$ is a normal subgroup of $E(2)$ and $O(2)$ is not

3.5 Lie groups

A **Lie group** is a differentiable manifold (something like a smooth surface in an arbitrary number of dimensions), where the group operation is an analytic map, as is the inversion operation that gives the inverse of a group element. Group elements can be varied continuously, so a Lie group is a way of describing a continuous symmetry.

Example 3.19 *The group of rotations in the plane, $SO(2)$, is a Lie group. A group element is described by an angle of rotation, θ. Since θ also describes a position on the unit circle, the group also has the structure of a manifold – a circle in this case. The angle of rotation, and hence the group elements, can be varied continuously. Composition of rotations is given by the map $f : SO(2) \times SO(2) \to SO(2)$, defined by*

$$f(\theta, \phi) = \theta + \phi, \quad \forall \theta, \phi \in SO(2), \tag{3.70}$$

which is clearly analytic. Similarly inversion is given by the map $i : SO(2) \to SO(2)$ defined by

$$i(\theta) = -\theta, \quad \forall \theta \in SO(2), \tag{3.71}$$

which again is analytic.

$SO(2)$ is a **one-parameter** *Lie group, where the rotation, θ, is the parameter.*

Example 3.20 *The general linear group, $GL(n)$, is also a Lie group. The space of all $n \times n$ matrices can be identified with $\mathbb{R}^{n^2}$, the space in which the n^2 real entries of the matrices are coordinates. $GL(n)$, the space of all $n \times n$ invertible matrices, is an open submanifold of this n^2-dimensional space.*

A Lie group is **compact** if its manifold is compact, in other words if it can be realised as a closed and bounded submanifold of $\mathbb{R}^n$ for some n. Equivalently, a Lie group is compact if its parameters vary over a closed interval. So, for example, $SO(2)$ is a compact Lie group because its elements are given by a parameter θ that describes a circle and lies in the range $0 \leq \theta \leq 2\pi$, which is bounded below by 0 and above by 2π.

A **closed subgroup** of $GL(n)$ is a closed subset of the manifold of $GL(n)$ that is also a subgroup of $GL(n)$. An **open subgroup** of $GL(n)$ is one that is not closed. It can be shown that every compact Lie group is isomorphic to a closed subgroup of $GL(n)$, the group of all invertible $n \times n$ matrices over $\mathbb{R}$. (See Bröcker and tom Dieck, 1985.)

Every finite group, Γ, is also isomorphic to a closed subgroup of $GL(n)$, since applying any element $\gamma \in \Gamma$ to Γ permutes the order of the group elements. If the permutation matrix corresponding to this permutation is P_γ, and Γ_P is the group of permutation matrices obtained in this way, then the isomorphism $\theta : \Gamma \to \Gamma_P$ is given by $\theta(\gamma) = P_\gamma$.

Many of the results we shall use in Chapters 4, 5 and 6 only hold for finite groups and compact Lie groups, so we shall mostly be dealing with groups that are isomorphic to closed subgroups of invertible real matrices.

3.6 Representations of groups

A **representation** (strictly a **linear representation**) of a finite group or compact Lie group, Γ, over a field, F, is a homomorphism θ from Γ to the group of matrices $GL(n, F)$. The **degree** or **dimension** of the representation is n.

The representation gives the **action** of Γ on the vector space $V = F^n$. We shall only consider the cases where F is $\mathbb{R}$ or $\mathbb{C}$, and the representations map to groups of real or complex matrices. The words representation and action are often used fairly interchangeably, and I shall follow suit.

More loosely we say that an n-dimensional representation of a group, Γ, is a set of invertible $n \times n$ matrices that conform to the group structure of Γ under matrix multiplication. If the representation associates a matrix, M_γ, with the group element γ, according to $\theta(\gamma) = M_\gamma$, then the set of matrices must satisfy

$$M_{\gamma_2} M_{\gamma_1} = M_{\gamma_2 \gamma_1}, \quad \forall \gamma_1, \ \gamma_2 \in \Gamma. \tag{3.72}$$

A group can have many different representations of various different dimensions Every representation has $M_e = I_n$, where e is the identity element.

Example 3.21 *The group D_3 has the representation*

$$M_e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad M_m = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},$$
$$M_\rho = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}, \quad M_{m\rho} = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix},$$
$$M_{\rho^2} = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}, \quad M_{m\rho^2} = \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}, \tag{3 73}$$

which is the set of matrices you would write down to describe the corresponding transformations of the plane in coordinate geometry This is sometimes referred to as the **natural** *representation D_3 also has other representations such as the identity representation that we will describe in the next example.*

Example 3.22 (The identity representation) *Every group Γ has a one-dimensional* **identity** *or* **trivial** *representation The identity representation is given by $M_\gamma = 1$, $\forall \gamma \in \Gamma$ Clearly this conforms to the group structure since $M_{\gamma_1} M_{\gamma_2} = 1 \times 1 = 1 = M_{\gamma_1 \gamma_2}$, $\forall \gamma_1, \gamma_2 \in \Gamma$*

A representation is said to be **faithful** if the mapping from the group elements to the matrices of the representation is an isomorphism The representation of D_3 in Example 3.21 is faithful as it is clearly bijective, but the identity representation is **unfaithful** since it maps every group element to 1

If two groups Γ and Δ are isomorphic, with the isomorphism $\phi : \Gamma \to \Delta$, and if the homomorphism $\theta : \Delta \to GL(n, F)$ is a representation of Δ, then the homomorphism $\theta\phi : \Gamma \to GL(n, F)$ is a representation of Γ. So every representation θ of Δ gives rise to a unique representation $\theta\phi$ of Γ that maps to the same group of matrices. Similarly every representation ψ of Γ gives rise to a unique representation $\psi\phi^{-1}$ of Δ that maps to the same group of matrices. Thus isomorphic groups have the same set of matrix representations.

Two representations are said to be **equivalent** or **isomorphic** if there is an invertible matrix S which maps between the matrices A_i of one representation and the matrices B_i of the other representation according to

$$A_i = S^{-1} B_i S, \quad \forall i \tag{3 74}$$

Equivalent representations have the same dimension This transformation by S is equivalent to changing the basis of the set of matrices, that is, making a coordinate transformation, so equivalent representations are just the same representation

in different coordinates. More abstractly, if Γ is a compact Lie group acting on two n-dimensional vector spaces V and W, the actions on V and W are said to be **isomorphic**, or equivalently V and W are **Γ-isomorphic**, if there is a linear isomorphism (a bijective linear map between vector spaces) $\theta : V \to W$ that commutes with the action of Γ:

$$\theta(\gamma v) = \gamma(\theta(v)), \quad \forall v \in V, \ \forall \gamma \in \Gamma \tag{3.75}$$

Here γv is shorthand for $\phi_1(\gamma)v$, where ϕ_1 is the representation of Γ on V. Similarly $\gamma(\theta(v))$ is shorthand for $\phi_2(\gamma)(\theta(v))$, where ϕ_2 is the representation of Γ on W. We shall use this kind of shorthand a lot in subsequent chapters for simplicity of notation.

Example 3.23 *Consider the representation of $\mathcal{D}_3$ given by $\rho : z \to ze^{2\pi i/3}$ and $m : z \to -\bar{z}$, where $z \in \mathbb{C}$ and where $\bar{z}$ is the complex conjugate of z. It is straightforward to work out the action of the other group elements by combining powers of the generators ρ and m, since the representation is a homomorphism and therefore must respect the group structure. We will now show that this representation is isomorphic to the natural representation. The isomorphism $\theta : \mathbb{C} \to \mathbb{R}^2$ is defined by*

$$\theta(x + iy) = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \forall x, y \in \mathbb{R} \tag{3.76}$$

We can check that the isomorphism commutes with $\mathcal{D}_3$ by considering the generators in turn. Take the action of ρ first. We have

$$\rho\theta(x + iy) = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -\frac{x}{2} - \frac{\sqrt{3}y}{2} \\ \frac{\sqrt{3}x}{2} - \frac{y}{2} \end{pmatrix} \tag{3.77}$$

and

$$\begin{aligned} \theta(\rho(x + iy)) &= \theta(e^{2\pi i/3}(x + iy)) \\ &= \theta\left(\left(-\frac{1}{2} + i\frac{\sqrt{3}}{2}\right)(x + iy)\right) \\ &= \theta\left(-\frac{x}{2} - \frac{\sqrt{3}y}{2} + i\frac{\sqrt{3}x}{2} - i\frac{y}{2}\right) \\ &= \begin{pmatrix} -\frac{x}{2} - \frac{\sqrt{3}y}{2} \\ \frac{\sqrt{3}x}{2} - \frac{y}{2} \end{pmatrix}, \end{aligned} \tag{3.78}$$

and so $\rho\theta(x+iy) = \theta(\rho(x+iy))$ *Similarly we have*

$$m\theta(x+iy) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -x \\ y \end{pmatrix} \tag{3.79}$$

and

$$\theta(m(x+iy)) = \theta(-x+iy) = \begin{pmatrix} -x \\ y \end{pmatrix}, \tag{3.80}$$

and so $m\theta(x+iy) = \theta(m(x+iy))$ *Thus the two representations are isomorphic*

A **unitary representation** is one that is made up entirely of unitary matrices. A **unitary** matrix is a matrix M such that

$$M^\dagger = M^{-1}, \tag{3.81}$$

where $\dagger$ represents the Hermitian conjugate, namely the complex conjugate of the transpose A unitary matrix has determinant ± 1 or $\pm i$, since

$$1 = \det I = \det(M^\dagger M) = \det M^\dagger \det M = |\det M|^2. \tag{3.82}$$

A real matrix satisfying equation (3.81) is said to be **orthogonal** (see Example 3.15). For example, the natural representation of $\mathcal{D}_3$ given in equation (3.73) is orthogonal

It can be shown that if Γ is a finite group or a compact Lie group then every representation of Γ over $\mathbb{C}$ is equivalent to a unitary representation The proof can be found in Cornwell (1984). In general we shall assume that our representations are unitary (or orthogonal if they are real).

Within the collection of all representations of a group, Γ, there are two classes of special ones: irreducible and absolutely irreducible representations. Absolutely irreducible real representations are particularly important in the theory of steady bifurcations in symmetric systems that is developed in Chapter 4.

A subspace W of V is **Γ-invariant** under the representation θ of a group Γ if

$$\theta(\gamma)w \in W, \quad \forall \gamma \in \Gamma, \ \forall w \in W \tag{3.83}$$

A representation or action of Γ is said to be **irreducible** if the only Γ-invariant subspaces are the origin, $\{\mathbf{0}\}$, and the whole space, V.

The shorthand **irrep** is often used to refer to irreducible representations

An action or representation of Γ is said to be **absolutely irreducible** if the only linear mappings that commute with the action of Γ on V are scalar multiples of the identity

For representations over $\mathbb{C}$, there is no distinction between irreducibility and absolute irreducibility, but real representations can be irreducible without being absolutely irreducible as is shown in Example 3.26 (see Golubitsky, Stewart &

Schaeffer, 1988, for further discussion of this point). Nonabsolutely irreducible real representations are important in the theory of Hopf bifurcations in symmetric systems (see Section 4.4).

Representations that are not irreducible are said to be **reducible**.

Example 3.24 *All one-dimensional representations are absolutely irreducible, as is the natural representation of $\mathcal{D}_3$ (equation 3.73).*

Example 3.25 *The representation on $\mathbb{R}^n$ given by $M_\gamma = I_n$, $\forall \gamma \in \Gamma$, is reducible for $n > 1$, since every subspace is Γ-invariant and if $n > 1$ it is possible to choose a subspace that is neither the origin nor the whole of $\mathbb{R}^n$.*

If a representation is reducible, then there is a Γ-invariant subspace $W \subset V$ that is neither $\{\mathbf{0}\}$ nor V. Its orthogonal complement, U, is defined by

$$U = \{u : u^\dagger w = 0, \forall w \in W\} \tag{3.84}$$

If M_γ is unitary (or orthogonal for a real representation) it is nonsingular, and so for all $w \in W$ there exists a $w' \in W$ such that $w = M_\gamma w'$, namely $w' = M_\gamma^{-1} w$. Thus

$$(M_\gamma u)^\dagger w = (M_\gamma u)^\dagger M_\gamma w' = u^\dagger M_\gamma^\dagger M_\gamma w' = u^\dagger w' = 0, \quad \forall w \in W, \tag{3.85}$$

and so the orthogonal complement U is also Γ-invariant. In what follows we shall use γw and γu as shorthand for $M_\gamma w$ and $M_\gamma u$ respectively. We can represent a general vector $v \in V$ as the sum of a vector w in W and a vector u in the complement, U. Now the linear projection $p : V \to W$ defined by

$$p(w + u) = w, \quad \forall w \in W, \ \forall u \in U, \tag{3.86}$$

commutes with Γ since

$$\begin{aligned} &p(\gamma(w + u)) = p(\gamma w + \gamma u) = p(\gamma w) = \gamma w = \gamma p(w + u), \\ &\forall w \in W, \ \forall u \in U, \end{aligned} \tag{3.87}$$

since $\gamma w \in W$ and $\gamma u \in U$. Thus no reducible action can be absolutely irreducible, and so absolute irreducibility implies irreducibility.

Each matrix M_{γ_i} in a reducible representation can be decomposed into the form

$$M_{\gamma_i} = \begin{pmatrix} A_i & B_i \\ 0 & C_i \end{pmatrix}, \tag{3.88}$$

where the matrix A_i acts on the nontrivial Γ-invariant subspace. If the representation is unitary (or orthogonal) then $B_i = 0$ for all i, since the complement of the Γ-invariant subspace is also Γ-invariant. The groups of matrices A_i and C_i

themselves form separate representations of Γ, since

$$\begin{pmatrix} A_i & B_i \\ 0 & C_i \end{pmatrix} \begin{pmatrix} A_j & B_j \\ 0 & C_j \end{pmatrix} = \begin{pmatrix} A_i A_j & A_i B_j + B_i C_j \\ 0 & C_i C_j \end{pmatrix}, \tag{3.89}$$

so the A_i and C_i will conform to the group structure because the M_{γ_i} do If these lower-dimensional representations are themselves reducible the process can be repeated, and so on until we are left with irreducible representations along the diagonal Irreducible representations thus form the basic building blocks of the reducible representations.

Example 3.26 *In the standard action of $SO(2)$ on $\mathbb{R}^2$, a rotation through an angle, θ, is represented by the rotation matrix*

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{3.90}$$

The only $SO(2)$-invariant subspaces are the origin and the whole of $\mathbb{R}^2$, so this action is irreducible. However, it is not absolutely irreducible since each matrix R_θ commutes with every other rotation matrix $R_\phi \in SO(2)$:

$$R_\theta R_\phi = R_\phi R_\theta = R_{\theta+\phi} \tag{3.91}$$

This gives an idea of why irreducibility and absolute irreducibility are the same thing for complex representations, but not for real ones. Only complex multiples of the identity, $Ae^{i\theta} I_n$, where $A \in \mathbb{R}$, commute with an (absolutely) irreducible complex action. If we consider the linear isomorphism $\alpha : \mathbb{C} \to \mathbb{R}^2$ defined by

$$\alpha(x + iy) = \begin{pmatrix} x \\ y \end{pmatrix}, \forall\, x,\ y \in \mathbb{R}, \tag{3.92}$$

then

$$\alpha(e^{i\theta}(x + iy)) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{3.93}$$

So commuting with a complex multiple of the identity in $\mathbb{C}$ is like commuting with a real multiple of R_θ in $\mathbb{R}^2$.

On the other hand, if we consider R_θ to act on $\mathbb{C}^2$ then the matrix has eigenvalues, $e^{\pm i\theta}$, and corresponding eigenvectors $\begin{pmatrix} 1 \\ \mp i \end{pmatrix}$ The eigenvectors do not depend on θ, so the spaces (a, ia) and $(a, -ia)$, $a \in \mathbb{R}$, are each invariant under the action of $SO(2)$ on $\mathbb{C}^2$ defined by R_θ, which is therefore reducible In a basis

given by the eigenvectors, the matrices of the complex representation take the form

$$\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}, \tag{3.94}$$

showing clearly that this reducible representation is formed from two one-dimensional irreducible complex representations where the action of the rotation is given by $e^{\pm i\theta}$

Now consider the standard action of $O(2)$ *on* $\mathbb{R}^2$*, generated by the rotations, represented by the matrices* R_θ*, given above, and a reflection,* m*, represented by the matrix*

$$M = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.95}$$

Again the only $SO(2)$*-invariant subspaces are* $\{\mathbf{0}\}$ *and* $\mathbb{R}^2$*, so the representation is irreducible. However if we now look for matrices*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{3.96}$$

that commute with M *then*

$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} -a & -b \\ c & d \end{pmatrix}, \tag{3.97}$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -a & b \\ -c & d \end{pmatrix}, \tag{3.98}$$

and so we must have $b = c = 0$ *If the matrix is also to commute with* R_θ *for all rotation angles,* θ*, then it must satisfy*

$$\begin{aligned} \begin{pmatrix} a\cos\theta & -d\sin\theta \\ a\sin\theta & d\cos\theta \end{pmatrix} &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} \\ &= \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \\ &= \begin{pmatrix} a\cos\theta & -a\sin\theta \\ d\sin\theta & d\cos\theta \end{pmatrix}, \end{aligned} \tag{3.99}$$

and so we must have $a = d$ *Thus the only matrices that commute with this representation of* $O(2)$ *are scalar multiples of the identity, and so the standard action of* $O(2)$ *on* $\mathbb{R}^2$ *is absolutely irreducible*

The theory of steady bifurcation with symmetry uses the absolutely irreducible real representations. We shall see shortly that it is possible to work out the set of

irreducible representations over $\mathbb{C}$, so it would be useful to know how to deduce the absolutely irreducible real representations from these

If the only complex matrices that commute with all the matrices, M_γ, of a representation are complex multiples of the identity, $(a + ib)I_n$, where $a, b \in \mathbb{R}$, then the only real matrices that commute with all the M_γ are real multiples of the identity, aI_n, for $a \in \mathbb{R}$ This means that if an n-dimensional irreducible representation of a group, Γ, over $\mathbb{C}$ turns out to have real matrices, M_γ, for all $\gamma \in \Gamma$, then this real representation is an n-dimensional absolutely irreducible representation over $\mathbb{R}$

Conversely, if all the M_γ are real and the only matrices that commute with them are real multiples of the identity, then if a complex matrix, C, commutes with all the M_γ, it must satisfy

$$CM_\gamma = M_\gamma C, \quad \forall \gamma \in \Gamma. \tag{3.100}$$

Now we can write $C = A + iB$, where A and B are real matrices, so

$$(A + iB)M_\gamma = M_\gamma(A + iB), \quad \forall \gamma \in \Gamma. \tag{3.101}$$

Taking real and imaginary parts gives $AM_\gamma = M_\gamma A$ and $BM_\gamma = M_\gamma B$, for all $\gamma \in \Gamma$ Thus A and B must be real multiples of the identity, so $A = aI_n$ and $B = bI_n$ for some $a, b \in \mathbb{R}$, and so $C = (a + ib)I_n$ is a complex multiple of the identity This shows that an n-dimensional absolutely irreducible representation over $\mathbb{R}$ is also an n-dimensional irreducible representation over $\mathbb{C}$

What all this tells us is that each absolutely irreducible real representation of a group, Γ, is equivalent to an irreducible representation of the same dimension over $\mathbb{C}$ Thus, once we have found all the irreps of Γ over $\mathbb{C}$, we can deduce the absolutely irreducible real representations by working out which ones can be realised over $\mathbb{R}$, in other words, which ones can be written as a set of real matrices (by a change of basis if necessary) There is a systematic method for working out which complex representations can be realised over $\mathbb{R}$, involving something called the Frobenius–Schur count of involutions (see James and Liebeck, 1993) For the groups that we shall come across in our study of pattern formation it will usually be straightforward to write down the real matrices of any such representation.

Theorem 3.1 (Orthogonality theorem for matrix representations) *Let the sets of matrices M_γ^p and M_γ^q belong to two unitary representations of a finite group Γ, where p and q label the representations, so that the representations are identical if $p = q$ and inequivalent if $p \neq q$ Then*

$$\frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} \overline{(M_\gamma^p)}_{jk} (M_\gamma^q)_{st} = \frac{1}{n_p} \delta_{pq} \delta_{js} \delta_{kt}, \tag{3.102}$$

where n_p is the dimension of M^p.

A version of this theorem also holds for compact Lie groups, with the sum being replaced by an integral. The proofs can be found in Cornwell (1984).

The theorem says that if you make a vector out of corresponding matrix entries of one representation for each group element $\gamma \in \Gamma$, for example the row 1, column 1 entries from each $M(\gamma)$,

$$(M_\gamma)_{11} = \begin{pmatrix} (M_{\gamma_1})_{11} \\ (M_{\gamma_2})_{11} \\ (M_{\gamma_3})_{11} \\ \\ (M_{\gamma_n})_{11} \end{pmatrix}, \tag{3.103}$$

where n is the dimension of the irrep, then all the possible vectors are orthogonal to each other, that is, the dot product of one vector with the complex conjugate of another is zero. Furthermore the vectors constructed from one representation are orthogonal to vectors constructed from an inequivalent representation. If you form the dot product of any one of these vectors with its complex conjugate the result is $|\Gamma|/n_p$ where n_p is the dimension of the representation from which the vector is constructed.

Example 3.27 *Let us consider the natural and identity representations of D_3. Both these representations are real, so there is no need to take the complex conjugate of a vector when forming the dot product. Taking the natural representation first, the vectors $M^a(\Gamma)_{ij}$ are given by*

$$\left(M^{\mathrm{n}}_\Gamma\right)_{11} = \begin{pmatrix} 1 \\ -\frac{1}{2} \\ -\frac{1}{2} \\ -1 \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \left(M^{\mathrm{n}}_\Gamma\right)_{21} = \begin{pmatrix} 0 \\ \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} \\ 0 \\ \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} \end{pmatrix},$$

$$\left(M^{\mathrm{n}}_\Gamma\right)_{12} = \begin{pmatrix} 0 \\ -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} \\ 0 \\ \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} \end{pmatrix}, \left(M^{\mathrm{n}}_\Gamma\right)_{22} = \begin{pmatrix} 1 \\ -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \\ -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}, \tag{3.104}$$

where the superscript n *denotes the natural representation, and the rows of the vectors range over the group elements in the order $\{e, \rho, \rho^2, m, m\rho, m\rho^2\}$. Any*

one of these vectors is orthogonal to any other. The dot product of any of the vectors with itself is 3, *which is equal to* $|\Gamma|/n_{\mathrm{n}}$, *where the order of the group is* $|\Gamma| = 6$, *and the dimension of the irrep is* $n_{\mathrm{n}} = 2$. *Only one vector can be formed from the identity representation, and it is*

$$\left(M_{\Gamma}^{i}\right)_{11} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \tag{3 105}$$

where the superscript i denotes the identity representation. This vector is orthogonal to any of the vectors formed from the natural representation, and when dotted with itself gives 6, *which is equal to* $|\Gamma|/n_i$, *where* $n_i = 1$ *is the dimension of the irrep*

3.7 Characters

The **character**, $\chi(M)$, of an $n \times n$ matrix M is its trace:

$$\chi(M) = \sum_{i=1}^{n} M_{ii}. \tag{3 106}$$

The characters of the component matrices are helpful for working out the irreducible representations of a group, as we shall see below and in Chapter 4. The identity element is always represented by I_n, so has character n

Many results in the theory of characters are derived for the case of representations over $\mathbb{C}$, so we shall assume that the vector space, V, is $\mathbb{C}^n$ in this section, and so the set of matrices of the representation form a subgroup of $GL(n, \mathbb{C})$. Recall that in this context irreducibility and absolute irreducibility are the same thing. As discussed above, the matrices of the representation can be real, even though we are working in $V = \mathbb{C}^n$. If all the matrices of a given representation turn out to be real then the representation is a real representation.

If two group elements h_1 and h_2 in Γ are conjugate, satisfying

$$h_1 = \gamma h_2 \gamma^{-1}, \tag{3 107}$$

for some $\gamma \in \Gamma$, then for a given representation, the characters of the two corresponding matrices M_1 and M_2 must be the same, since

$$\begin{aligned}\chi(M_1) &= \sum_{i=1}^{n}(M_1)_{ii} = \sum_{i=1}^{n}\left(M_\gamma M_2 M_\gamma^{-1}\right)_{ii} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}(M_\gamma)_{ij}(M_2)_{jk}\left(M_\gamma^{-1}\right)_{ki} \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n}(M_\gamma^{-1}M_\gamma)_{kj}(M_2)_{jk} = \sum_{j=1}^{n}\sum_{k=1}^{n}\delta_{kj}(M_2)_{jk} \\ &= \sum_{j=1}^{n}(M_2)_{jj} = \chi(M_2), \end{aligned} \tag{3.108}$$

where M_γ is the matrix representing γ, and n is the dimension of the representation.

Example 3.28 *The characters of the natural representation of $\mathcal{D}_3$ are*

$$\chi_e = 2,\ \chi_\rho = -1,\ \chi_{\rho^2} = -1,\ \chi_m = 0,\ \chi_{m\rho} = 0,\ \chi_{m\rho^2} = 0. \tag{3.109}$$

The identity has character 2 as this is a two-dimensional irrep. All elements in the conjugacy class $\{m, m\rho, m\rho^2\}$ have character 0, while the two conjugate elements ρ and ρ^2 have character -1

The characters of a real representation, such as the one in Example 3.28, are real, but the converse is not necessarily true: in other words, a representation with all real characters need not necessarily be realisable over $\mathbb{R}$. For further discussion of this point, see James and Liebeck (1993).

For finite groups or compact Lie groups the characters of the irreducible representations specify the irreps up to equivalence. If $M^p(\gamma)$ and $M^q(\gamma)$ represent the group element, γ, in two n-dimensional equivalent representations, then by equivalence there exists a matrix S such that

$$M^p(\gamma) = SM^q(\gamma)S^{-1}, \quad \forall\gamma \in \Gamma, \tag{3.110}$$

and so following the argument in equation (3.108) above we have

$$\chi^p(\gamma) = \chi^q(\gamma), \quad \forall\gamma \in \Gamma, \tag{3.111}$$

where $\chi^p(\gamma)$ and $\chi^q(\gamma)$ are the characters of $M^p(\gamma)$ and $M^q(\gamma)$ respectively. Thus any two equivalent representations of Γ will have the same character systems, in other words each group element will have the same character in the two representations

Table 3.3 *A character table for a group, Γ, with conjugacy classes CS_j, irreps i and characters χ_{ij}, where $i = 1, \ldots, n$ and $j = 1, \ldots, n$*

Irrep	CS_1	CS_2	.	CS_n
1	χ_{11}	χ_{12}	...	χ_{1n}
2	χ_{21}	.		.
n	χ_{n1}	.		χ_{nn}

It is also possible to show that for a finite group or compact Lie group, equality of character systems is sufficient to prove equivalence of representations. The proof of this is given in Cornwell (1984). In principle then, if we can determine the characters, we should be able to work out the irreps.

The characters of the inequivalent irreps of a finite group Γ can be presented in a **character table** such as the one given in Table 3.3. The conjugacy classes are given across the top, the inequivalent irreps down the side of the table and the corresponding entries are the characters. Remember that all elements in a class have the same character.

Theorem 3.2 *The number of inequivalent irreducible representations of a finite group Γ is equal to the number of conjugacy classes of Γ.*

This means that character tables are always square as shown in Table 3.3.

Theorem 3.3 *The sum of the squares of the dimensions d_i of the n inequivalent irreducible representations of a finite group Γ, is equal to the order, $|\Gamma|$, of Γ:*

$$\sum_{i=1}^{n} d_i^2 = |\Gamma|. \tag{3.112}$$

The proofs of Theorems 3.2 and 3.3 can be found in Cornwell (1984).

Theorem 3.4 (Orthogonality theorems for characters) *For a finite group, Γ, the characters satisfy*

$$\sum_{p=1}^{n} \overline{\chi^p(CS_i)}\chi^p(CS_j)N_i = |\Gamma|\delta_{ij}, \tag{3.113}$$

where the sum is taken over all inequivalent irreducible representations of Γ, *and*

$$\sum_{i=1}^{m} \overline{\chi^p(CS_i)}\chi^q(CS_i)N_i = |\Gamma|\delta_{pq}, \tag{3.114}$$

where the sum is over the classes of Γ. *In both cases,* $\chi^p(CS_i)$ *is the character of the class* CS_i *in the representation* p *of* Γ, $|\Gamma|$ *is the order of* Γ *and* N_i *is the number of elements in the class* CS_i.

The proof of equation (3.113) is given in Cornwell (1984), while equation (3.114) can be obtained from the orthogonality theorem for matrices (Theorem 3.1) by setting $k = j$ and $t = s$ and summing over j and s.

The orthogonality theorems say that the columns of the character table are orthogonal and so are the rows when weighted by the size of the classes. The weighted sum of the squared moduli of the entries in each row or column adds up to the order of the group.

For finite groups of low order Theorems 3.2, 3.3 and 3.4 are often sufficient to determine the character table of the group, which we will use to deduce the irreps.

Example 3.29 *$\mathcal{D}_3$ has three conjugacy classes, so by Theorem 3.2 it must have three inequivalent irreps. By Theorem 3.3 the sum of the squares of the dimensions of the irreps must equal the order of the group, which is 6. That means that it must have two one-dimensional irreps and one two-dimensional irrep, since* $1^2 + 1^2 + 2^2 = 6$, *and there is no other way to add three square numbers to get 6. We already know about two of the irreps: the identity and natural representations, so we just need to find one more one-dimensional irrep, which turns out to be*

$$M^a_e = M^a_\rho = M^a_{\rho^2} = 1, \quad M^a_m = M^a_{m\rho} = M^a_{m\rho^2} = -1, \tag{3.115}$$

where a labels the irrep. In this irrep rotations and the identity are represented by $+1$ and reflections by -1, values equal to the determinant of the corresponding matrix in the natural representation. There is always a one-dimensional irrep of this kind for $\mathcal{D}_n$.

The classes of $\mathcal{D}_3$ are $\{e\}$, $\{\rho, \rho^2\}$ and $\{m, m\rho, m\rho^2\}$, and so we can construct the character table for $\mathcal{D}_3$ (Table 3.4). You can see that different columns are orthogonal to each other, as are the rows when weighted by the number of elements in each class. The dot product of each row or column with itself, weighted by the number of elements in each class, is 6, the order of the group.

Now we need to construct the matrices of the irreps from their characters. Although the sets of characters are unique, the corresponding sets of matrices are not, because equivalent representations have the same characters. So all we have

Table 3.4. *The character table for D_3. The labels i and n denote the identity and natural representations respectively, and the label a denotes the irrep defined in equation (3.115)*

Irrep	$\{e\}$	$\{\rho, \rho^2\}$	$\{m, m\rho, m\rho^2\}$
i	1	1	1
a	1	1	−1
n	2	−1	0

to do is find one possible irreducible representation that has the given character system.

In the case of one-dimensional irreps, the matrices are identical to the characters. For higher-dimensional irreps, we have to construct a set of matrices, with traces given by the characters, that form a representation that cannot be reduced into combinations of lower-dimensional irreps. Often, for the groups that we will look at, one of the higher-dimensional irreps, if there are any, will be the natural representation, which is easy to write down using coordinate geometry. The natural representation is not always irreducible, but it is easy to check this once you have the matrices. If the natural representation is not an irrep or if there is more than one irrep of dimension greater than one, the most practical method of finding the matrices for relatively small groups is by trial and error. The orthogonality theorem for representations (Theorem 3.1) may also be helpful. The character table tells you when you have found all the irreps. For finite groups of coordinate transformations in $\mathbb{R}^3$ there is a systematic procedure involving projection operators (see Cornwell 1984).

For finite Abelian groups, finding the set of irreducible representations is easy. Any two elements, γ_1 and γ_2, of an Abelian group Γ commute, satisfying $\gamma_1\gamma_2 = \gamma_2\gamma_1$, and therefore so must the matrices of the representation, so that

$$M_{\gamma_1} M_{\gamma_2} = M_{\gamma_2} M_{\gamma_1}, \quad \forall \gamma_1, \gamma_2 \in \Gamma. \tag{3.116}$$

So each M_γ commutes with all the matrices of the representation. Now if the representation is irreducible, the only linear mappings that commute with it are scalar multiples of the identity and so we must have

$$M_\gamma = c(\gamma) I, \quad \forall \gamma \in \Gamma \tag{3.117}$$

for some suitable complex scalars $c(\gamma)$. Such a representation can only be irreducible if it is one-dimensional, and further, we can choose it to be unitary. If the

group has N generators, γ_j, each of order n_j, then we have $M^{(p)}_{\gamma_j^{n_j}} = 1$, and so

$$M^{(p)}_{\gamma_j^m} = e^{2\pi i p m/n_j}, \tag{3.118}$$

for $m = 0, \ldots, n_j - 1$, where $p = 1, \ldots, n_j$ labels the n_j inequivalent irreducible representations of the cyclic group generated by γ_j. The $\prod_{j=1}^{N} n_j$ inequivalent irreducible representations of the whole group can be constructed using all the possible combinations of the irreps of the cyclic groups.

For an Abelian group, Γ, we have

$$\gamma h \gamma^{-1} = \gamma\gamma^{-1} h = h \quad \forall h, \gamma \in \Gamma, \tag{3.119}$$

so each group element is in a class on its own. There are $\prod_{j=1}^{N} n_j$ group elements, $\prod_{j=1}^{N} \gamma_j^{m_j}$ for $m_j = 0, \ldots, n_j - 1$, and so there must be $\prod_{j=1}^{N} n_j$ conjugacy classes. Thus, as expected, the number of irreps is the same as the number of conjugacy classes.

3.8 Isotypic decomposition

Under the action of a compact Lie group, Γ, a vector space, V, can be decomposed into the sum of a finite number, m, of Γ-irreducible subspaces, U_i, giving

$$V = U_1 \oplus U_2 \oplus \ldots \oplus U_m. \tag{3.120}$$

This decomposition is not unique. Some of the U_i will be Γ-isomorphic to each other. Recall that two subspaces U_j and U_k are Γ-isomorphic if there is a linear isomorphism $\theta : U_j \to U_k$ that commutes with the action of Γ:

$$\theta(\gamma \boldsymbol{u}_j) = \gamma(\theta(\boldsymbol{u}_j)), \quad \forall \boldsymbol{u}_j \in U_j, \forall \gamma \in \Gamma. \tag{3.121}$$

Summing each set of Γ-isomorphic subspaces will give a subspace W_j, and these W_j are the **isotypic components** of V.

Conversely, each isotypic component, W, can be written as the direct sum of a set of isomorphic Γ-irreducible subspaces, giving

$$W = U^{(1)} \oplus U^{(2)} \ldots \oplus U^{(n)}, \tag{3.122}$$

where the superscripts label the subspaces, which are not necessarily the same as any of the U_1 to U_m above, and where $U^{(2)}, \ldots U^{(n)}$ are all Γ-isomorphic to $U^{(1)}$.

The space V can then be written as the direct sum of the isotypic components to give

$$V = W_1 \oplus W_2 \oplus \ldots \oplus W_k, \quad k \leq m. \tag{3.123}$$

This **isotypic decomposition** is unique, and is helpful for working out the stability of solutions to Γ-symmetric bifurcation problems on V, as we will see in Chapter 4. Proofs of the existence and uniqueness of the isotypic decomposition, can be found in Golubitsky, Stewart and Schaeffer (1988).

Example 3.30 *Let $SO(2)$ act by matrix multiplication on the space, V, of 2×2 real matrices, such that a rotation, ρ, through an angle θ is given by*

$$\rho \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{3.124}$$

Any 2×2 real matrix can be expressed as the sum of two other matrices in the manner

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & 0 \\ c & 0 \end{pmatrix} + \begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix}, \tag{3.125}$$

and so we can write $V = V_1 \oplus V_2$ where V_1 is the space of all real matrices of the form

$$\begin{pmatrix} a & 0 \\ c & 0 \end{pmatrix} \tag{3.126}$$

and V_2 is the space of all real matrices of the form

$$\begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix} \tag{3.127}$$

Now the action of $SO(2)$ on V_1 and V_2 is isomorphic to its standard irreducible matrix-multiplication action on $\mathbb{R}^2$ via the isomorphisms $\theta_1 : V_1 \to \mathbb{R}^2$ and $\theta_2 : V_2 \to \mathbb{R}^2$ defined by

$$\theta_1 \begin{pmatrix} a & 0 \\ c & 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix}, \tag{3.128}$$

$$\theta_2 \begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix}, \tag{3.129}$$

and hence V_1 and V_2 are $SO(2)$-isomorphic. The isotypic components are given by summing the $SO(2)$-isomorphic subspaces. Thus in this case the single isotypic component is the sum of V_1 and V_2, in other words the whole space, V. Clearly the isotypic decomposition is unique as it consists of one component: the whole space. On the other hand, the decomposition of V into irreducible subspaces is not unique. $SO(2)$ also acts irreducibly on the subspace V_3 consisting of all matrices

of the form

$$\begin{pmatrix} b & b \\ d & d \end{pmatrix}, \tag{3.130}$$

so we could write $V = V_1 \oplus V_3$ *instead of* $V = V_1 \oplus V_2$

Exercises

3.1 Show by checking the group axioms directly, that the integers, 0, 1, . . . , $n-1$, form a group under addition modulo n.

3.2 Show, using a group table, that the symmetry group of a square, $\mathcal{D}_4$, satisfies the group axioms

3.3 Show, using diagrams, that the symmetry group of a rectangle, $\mathcal{D}_2$, is Abelian.

3.4 Find the subgroups of $\mathcal{D}_4$ and for each nontrivial subgroup, work out whether it is normal

3.5 What are the normalizers of the subgroups $\mathcal{D}_2$ and $\mathbb{Z}_2$ (where $\mathbb{Z}_2$ is generated by a reflection) of $\mathcal{D}_6$?

3.6 Find the conjugacy classes of $\mathcal{D}_6$.

3.7 Find the irreducible representations of $\mathcal{D}_2$ over $\mathbb{C}$

3.8 Find the irreducible representations of C_3 over $\mathbb{C}$ Deduce a real nonabsolutely irreducible representation

3.9 If the action of $\mathbb{Z}_2$ on $\mathbb{R}_2$ is given by

$$M_e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad M_m = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},$$

find the isotypic decomposition

4

Bifurcations with symmetry

Group theory turns out to be very useful for analysing bifurcations in systems with symmetry, including those where patterns are seen. The application of group theory to bifurcations with symmetry is known as equivariant bifurcation theory. It is a rich and well-developed field and in this chapter we will just skim the surface, introducing a few of the most useful results. A trio of good books that go into a great deal more depth are those by Golubitsky, Stewart and Schaeffer (1988), Chossat and Lauterbach (2000) and Golubitsky and Stewart (2002). Sections 4.1, 4.2 and 4.4 follow the treatment given in Golubitsky, Stewart and Schaeffer, though abbreviated.

We are going to look at how the symmetry of observed patterns is affected or determined by the symmetry of the experimental set-up or governing equations that produce them. Before we can see how this works in practice, we need to spend the next two sections defining some useful concepts associated with symmetric equations and solutions, and introducing the equivariant branching lemma, the main result that relates the symmetry of steady solutions to the symmetry of the pattern-forming system in which they are seen.

Throughout this and subsequent chapters if $\gamma \in \Gamma$ is a group element and $\boldsymbol{x} \in V$ is a vector, then $\gamma\boldsymbol{x}$ is shorthand for $\theta(\gamma)\boldsymbol{x}$ where θ is the representation or action of Γ on V.

4.1 Ordinary differential equations with spatial symmetry

Consider the equation

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = f(\boldsymbol{x}, \mu), \tag{4.1}$$

where t is time, $\mu \in \mathbb{R}$ is a bifurcation parameter and $\boldsymbol{x} \in V$ is the position in phase space. We will work with real vector spaces, V, but it will often be convenient to

use complex notation to describe them, in which case we make the identification $\mathbb{R}^{2n} \cong \mathbb{C}^n$. We are interested in the situation where there is a group, Γ, of spatial symmetries with a (linear) matrix representation. If we apply one of the symmetries, $\gamma \in \Gamma$, to $\boldsymbol{x}$ to get $\boldsymbol{y} = \gamma \boldsymbol{x}$, then since γ is a symmetry of the system we want the equation for $\boldsymbol{y}$ to be the same as the one for $\boldsymbol{x}$, in other words

$$\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t} = f(\boldsymbol{y}, \mu) = f(\gamma \boldsymbol{x}, \mu), \tag{4.2}$$

but since γ is a purely spatial linear action, we also have

$$\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t} = \gamma \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \gamma f(\boldsymbol{x}, \mu) \tag{4.3}$$

Equating the two expressions for $\mathrm{d}\boldsymbol{y}/\mathrm{d}t$ from equations (4.2) and (4.3), we find that $f(\boldsymbol{x}, \mu)$ must satisfy

$$\gamma f(\boldsymbol{x}, \mu) = f(\gamma \boldsymbol{x}, \mu), \quad \forall \gamma \in \Gamma, \tag{4.4}$$

which implies a restriction on the form of $f(\boldsymbol{x}, \mu)$. Equation (4.4) is said to express an **equivariance condition** on $f(\boldsymbol{x}, \mu)$. Equivariance is the requirement that the left- and right-hand sides of equation (4.1) transform in a compatible way under the symmetries of the group. Note that the symmetries γ act only on the phase-space vector, $\boldsymbol{x}$, and not on the bifurcation parameter, μ.

We can use the equivariance condition to deduce the most general form of $f(\boldsymbol{x}, \mu)$, bearing in mind that the choice of representation will affect the outcome. If the matrix M_γ represents the group element γ, then the equivariance condition can be written

$$M_\gamma f(\boldsymbol{x}, \mu) = f(M_\gamma \boldsymbol{x}, \mu), \quad \forall \gamma \in \Gamma. \tag{4.5}$$

It is clear from this that the choice of the matrices M_γ will affect the form taken by $f(\boldsymbol{x}, \mu)$. For example if M_γ is $-I$ then f must be odd, satisfying $-f(\boldsymbol{x}, \mu) = f(-\boldsymbol{x}, \mu)$.

Equation (4.1) describes a **steady-state bifurcation problem** if it has a fixed point $\boldsymbol{x}_0$ such that $Df\big|_{(\boldsymbol{x}_0, \mu)}$ has a real eigenvalue passing through zero at the bifurcation point, $\mu = \mu_c$. From now on we shall assume that for any steady-state bifurcation problem we have chosen the origins of $\boldsymbol{x}$ and μ so that the fixed point is at $\boldsymbol{x} = \boldsymbol{0}$ and the bifurcation point is at $\mu = 0$. Then the existence of the fixed point at the bifurcation point requires that $f(\boldsymbol{0}, 0) = \boldsymbol{0}$. We shall also assume that we have reduced the system to the centre manifold so that the Jacobian vanishes at the bifurcation point, that is $Df\big|_{(\boldsymbol{0}\,0)} = 0$. It will be useful for later to note that if

Df_μ, defined by

$$(Df_\mu)_{ij} = \frac{\partial (Df)_{ij}}{\partial \mu}, \tag{4.6}$$

has nonzero determinant at $x = 0$, $\mu = 0$ then the eigenvalue passes through zero with nonzero speed

4.1.1 Group orbits

If x is an equilibrium solution of equation (4.1) then so is γx for all symmetries, γ, in the group because

$$\frac{d(\gamma x)}{dt} = f(\gamma x, \mu) = \gamma f(x, \mu) = 0. \tag{4.7}$$

The point γx is said to lie in the **orbit** of the action of Γ on $x \in V$ defined by

$$\Gamma x = \{\gamma x : \gamma \in \Gamma\} \tag{4.8}$$

The point x need not be a stationary solution of equation (4.1) in order for its orbit to be defined. However all stationary solutions on the same orbit have the same existence properties

Differentiating the equivariance condition (4.4) gives

$$\gamma Df\big|_{(x,\mu)} = Df\big|_{(\gamma x,\mu)}\gamma, \quad \forall \gamma \in \Gamma \tag{4.9}$$

If v is an eigenvector of $Df\big|_{(x,\mu)}$ with eigenvalue λ so that

$$Df\big|_{(x,\mu)} v = \lambda v, \tag{4.10}$$

then we also have

$$Df\big|_{(\gamma x,\mu)} \gamma v = \gamma Df\big|_{(x,\mu)} v = \gamma \lambda v = \lambda \gamma v, \tag{4.11}$$

using the linearity of the action of Γ on V. So γv is an eigenvector of $Df\big|_{(\gamma x,\mu)}$ with the same eigenvalue λ. Thus the eigenvalues of $Df\big|_{(x,\mu)}$ and $Df\big|_{(\gamma x,\mu)}$ are the same. If x and γx are fixed points on the same orbit of Γ they will have the same stability properties as well as the same existence properties, so they are the same type of solution

Equation (4.9) also tells us to expect multiple zero eigenvalues at a bifurcation with symmetry, because if $Df\big|_{(0,\mu)} v = 0$ then $Df\big|_{(0,\mu)} \gamma v = 0$ too. In other words, if v is an eigenvector with zero eigenvalue then so too is any point γv in the orbit of v.

If $\gamma \in \Gamma$ is a continuous symmetry depending smoothly on some parameter, θ, such that $\gamma \equiv \gamma(\theta)$, with $\gamma(0) = e$, then if x is a stationary point of equation (4.1),

so is $\gamma(\theta)\boldsymbol{x}$, for all θ, and so we have

$$f(\gamma\boldsymbol{x}, \mu) = \mathbf{0} \tag{4.12}$$

Differentiating with respect to θ and evaluating at $\theta = 0$ gives

$$Df\big|_{(\boldsymbol{x},\mu)} \left.\frac{\mathrm{d}\gamma}{\mathrm{d}\theta}\right|_{\theta=0} \boldsymbol{x} = \mathbf{0}, \tag{4.13}$$

and so $Df\big|_{(\boldsymbol{x},\mu)}$ has a zero eigenvalue with eigenvector

$$\left.\frac{\mathrm{d}\gamma}{\mathrm{d}\theta}\right|_{\theta=0} \boldsymbol{x}, \tag{4.14}$$

which is tangent to the group orbit, $\Gamma\boldsymbol{x}$. This means that the fixed point $\boldsymbol{x}$ has a zero growth rate eigenvalue corresponding to perturbations along the group orbit. We will see this sort of thing often with translation symmetries.

4.1.2 Isotropy subgroups

Now we want to know what the symmetry of the bifurcation problem (4.1) can tell us about the bifurcating solutions. To do this we look at the symmetry of the stationary solutions themselves.

The symmetry of a stationary solution, $\boldsymbol{x}$, of (4.1) is specified by its **isotropy subgroup**, $\Sigma_{\boldsymbol{x}} \subset \Gamma$, defined by

$$\Sigma_{\boldsymbol{x}} = \{\sigma \in \Gamma : \sigma\boldsymbol{x} = \boldsymbol{x}\} \tag{4.15}$$

So for a subgroup to be an isotropy subgroup it must fix some vector $\boldsymbol{x}$, and must contain all the group elements that fix $\boldsymbol{x}$. Whether a given subgroup of Γ is an isotropy subgroup will depend on the representation or action of Γ. An isotropy subgroup is closed as a natural consequence of its definition by equation (4.15).

Lemma 4.1 *Points on the same orbit of Γ have conjugate isotropy subgroups satisfying*

$$\Sigma_{\gamma\boldsymbol{x}} = \gamma\Sigma_{\boldsymbol{x}}\gamma^{-1} \tag{4.16}$$

Proof First we show that all elements of $\gamma\Sigma_{\boldsymbol{x}}\gamma^{-1}$ fix $\gamma\boldsymbol{x}$. We have

$$\gamma\sigma\gamma^{-1}(\gamma\boldsymbol{x}) = \gamma\sigma\gamma^{-1}\gamma\boldsymbol{x} = \gamma\sigma\boldsymbol{x} = \gamma\boldsymbol{x}, \quad \forall\gamma \in \Gamma,\ \forall\sigma \in \Sigma_{\boldsymbol{x}} \tag{4.17}$$

which shows that $\gamma\sigma\gamma^{-1} \in \Sigma_{\gamma\boldsymbol{x}}\ \forall\gamma \in \Gamma, \sigma \in \Sigma_{\boldsymbol{x}}$, and hence $\gamma\Sigma_{\boldsymbol{x}}\gamma^{-1} \subset \Sigma_{\gamma\boldsymbol{x}}$. Second we show that all elements of $\gamma^{-1}\Sigma_{\gamma\boldsymbol{x}}\gamma$ fix $\boldsymbol{x}$. We have

$$\gamma^{-1}\sigma\gamma\boldsymbol{x} = \gamma^{-1}\gamma\boldsymbol{x} = \boldsymbol{x}, \quad \forall\gamma \in \Gamma,\ \forall\sigma \in \Sigma_{\gamma\boldsymbol{x}} \tag{4.18}$$

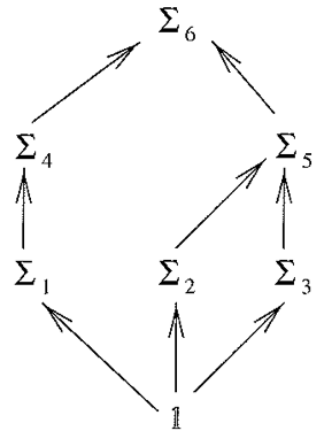


Fig 4.1 An example of an isotropy lattice. Inclusion is indicated by an arrow

which shows that $\gamma^{-1}\sigma\gamma \in \Sigma_x$, $\forall\gamma \in \Gamma$, $\forall\sigma \in \Sigma_{\gamma x}$, and hence $\gamma^{-1}\Sigma_{\gamma x}\gamma \subset \Sigma_x$ or equivalently $\Sigma_{\gamma x} \subset \gamma\Sigma_x\gamma^{-1}$ Since the inclusion holds both ways round we must have $\Sigma_{\gamma x} = \gamma\Sigma_x\gamma^{-1}$. □

The **conjugacy class** of an isotropy subgroup, Σ_x, is the set of all isotropy subgroups that are conjugate to Σ_x Each class is labelled by, or named after, one of its members, so we might refer to the conjugacy class Σ_x or $\mathbb{Z}_2$, for example. The isotropy subgroups of all points on one group orbit of Γ are in the same conjugacy class Since all points on the same group orbit have the same existence and stability properties, we regard conjugate isotropy subgroups as 'the same' in some loose sense.

An **orbit type**, W_x, of the action of Γ on V is the set of all points of V that have isotropy subgroups conjugate to Σ_x:

$$W_x = \{y \in V : \Sigma_y = \gamma\Sigma_x\gamma^{-1} \text{ for some } \gamma \in \Gamma\}. \tag{4.19}$$

The **isotropy lattice**, or **lattice of isotropy subgroups**, of a particular action of Γ is the set of all conjugacy classes of isotropy subgroups of Γ, partially ordered by inclusion. If Σ and Δ are distinct conjugacy classes of isotropy subgroups, then Σ is included in Δ if and only if there are isotropy subgroups $\Sigma_i \in \Sigma$ and $\Delta_j \in \Delta$ such that $\Sigma_i \subseteq \Delta_j$ Isotropy lattices are generally presented in the form of a diagram, such as the one in Figure 4.1 Arrows indicate inclusion: the class at the tail of the arrow is included in the class at the head

The Jacobian matrix, Df, evaluated at a stationary solution, x, of equation (4.1) commutes with all the elements of the isotropy subgroup, Σ_x, of x. You can see

this by putting $\gamma = \sigma \in \Sigma_x$ in equation (4.9) to give

$$\sigma Df\big|_{(x\ \mu)} = Df\big|_{(\sigma x\ \mu)}\sigma = Df\big|_{(x\ \mu)}\sigma, \quad \forall \sigma \in \Sigma_x, \tag{4.20}$$

since $\sigma x = x$, $\forall \sigma \in \Sigma_x$. A consequence of this is that $Ker(Df)$ is Σ_x-invariant, since

$$Df\sigma y = \sigma Df y = \mathbf{0}, \quad \forall y \in Ker(Df), \tag{4.21}$$

where from now on Df represents $Df\big|_{(x,\mu)}$. Recall that $Ker(\alpha)$ is the set of all points $x \in V$ such that $\alpha x = \mathbf{0}$, where α is a linear map acting on the vector space, V.

Now we are going to show that the isotypic decomposition of $\mathbb{R}^n$ with respect to the action of Σ_x, introduced in Section 3.8, block diagonalises Df, which is useful for computing the stability of the solution x. If

$$W = U^{(1)} \oplus U^{(2)} \ldots \oplus U^{(k)} \tag{4.22}$$

is an isotypic component, with the $U^{(i)}$ all Σ_x-irreducible, and $U^{(2)}, \ldots, U^{(k)}$ all isomorphic to $U^{(1)}$, then $Ker(Df|U^{(i)})$ is Σ_x-invariant for any $U^{(i)}$. Since Σ_x acts irreducibly on $U^{(i)}$, the only Σ_x-invariant subspaces can be $\{\mathbf{0}\}$ or the whole of $U^{(i)}$, so $Ker(Df|U^{(i)})$ is one of these. If $Ker(Df|U^{(i)}) = U^{(i)}$, then $Df|U^{(i)} = 0$, and so $Df(U^{(i)}) \subset W$. On the other hand if $Ker(Df|U^{(i)}) = \{\mathbf{0}\}$, then $Df|U^{(i)}$ is a linear isomorphism that commutes with the action of Σ_x, so $Df(U^{(i)}) \cong U^{(i)} \cong U^{(1)}$. Any Σ_x-irreducible subspace that is isomorphic to $U^{(1)}$ is contained in W, by the definition of an isotypic component, so we have $Df(U^{(i)}) \subseteq W$, for all $U^{(i)}$, and since W is a direct sum of the $U^{(i)}$ we have $Df(W) \subseteq W$. So each isotypic component is invariant under Df. In other words, the isotypic decomposition with respect to Σ_x block diagonalises Df.

4.1.3 Fixed-point subspaces

The **fixed-point subspace**, $\mathrm{Fix}(\Sigma)$, associated with a subgroup, Σ, of Γ is defined by

$$\mathrm{Fix}(\Sigma) = \{x \in V : \sigma x = x, \forall \sigma \in \Sigma\}. \tag{4.23}$$

Since the action of σ is linear, we have $\sigma\mathbf{0} = \mathbf{0}$ and so $\mathrm{Fix}(\Sigma)$ will always contain the origin and cannot be empty. As $\mathrm{Fix}(\Sigma)$ is not empty, then it is a subspace of V if it is closed under addition and scalar multiplication. If x and y are in $\mathrm{Fix}(\Sigma)$, then so is $x + y$ since

$$\sigma(x + y) = \sigma x + \sigma y = x + y, \quad \forall \sigma \in \Sigma, \tag{4.24}$$

since the action of σ is linear So $\mathrm{Fix}(\Sigma)$ is closed under addition. Similarly if $\boldsymbol{x}$ is in $\mathrm{Fix}(\Sigma)$, then so is $c\boldsymbol{x}$ for any constant c, because all linear actions, σ, commute with scalar multiples of the identity and so

$$\sigma c\boldsymbol{x} = c\sigma\boldsymbol{x} = c\boldsymbol{x}, \quad \forall \sigma \in \Sigma. \tag{4.25}$$

So $\mathrm{Fix}(\Sigma)$ is also closed under scalar multiplication, and hence it is indeed a subspace of V. Remember that for any given group the fixed-point subspaces will depend upon the choice of representation

Fixed-point subspaces are flow-invariant since

$$f(\boldsymbol{x}, \mu) = f(\sigma\boldsymbol{x}, \mu) = \sigma f(\boldsymbol{x}, \mu), \quad \forall \boldsymbol{x} \in \mathrm{Fix}(\Sigma). \tag{4.26}$$

This says that $f(\mathrm{Fix}(\Sigma), \mu) \subset \mathrm{Fix}(\Sigma)$. So if we know we are looking for a solution to

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = f(\boldsymbol{x}, \mu), \tag{4.27}$$

with a certain isotropy subgroup, $\Sigma_{\boldsymbol{x}}$, we can just restrict f to $\mathrm{Fix}(\Sigma_{\boldsymbol{x}})$ and solve the equation there

The two trivial subgroups of Γ are the whole group Γ, and the identity subgroup, $\mathbb{1} = \{e\}$, which fixes every point $\boldsymbol{x} \in V$, so that $\mathrm{Fix}(\mathbb{1}) = V$. $\mathrm{Fix}(\Gamma)$ consists of all the points $\boldsymbol{x} \in V$ that are fixed by every group element. It will often be useful to consider the case where only the origin is fixed by the whole group, so that $\mathrm{Fix}(\Gamma) = \{\mathbf{0}\}$. This is not always true, but does hold as long as we use nontrivial irreducible representations It is easy to see that $\mathrm{Fix}(\Gamma)$ is Γ-invariant since

$$\gamma\boldsymbol{x} = \boldsymbol{x}, \ \forall \boldsymbol{x} \in \mathrm{Fix}(\Gamma) \tag{4.28}$$

By definition if Γ acts irreducibly, the only Γ-invariant subspaces of V are $\{\mathbf{0}\}$ and V So either $\mathrm{Fix}(\Gamma) = \{\mathbf{0}\}$ or $\mathrm{Fix}(\Gamma) = V$, but if $\mathrm{Fix}(\Gamma) = V$ then Γ acts trivially, so we are left with $\mathrm{Fix}(\Gamma) = \{\mathbf{0}\}$

If $\mathrm{Fix}(\Gamma) = \{\mathbf{0}\}$ then equation (4.26) says that there is a trivial solution to equation (4 1) since then $f(\mathbf{0}, \mu) \in \mathrm{Fix}(\Gamma)$ and so $f(\mathbf{0}, \mu) = \mathbf{0}$.

Another useful result is that if $\mathrm{Fix}(\Gamma) = \{\mathbf{0}\}$ then $\dim \mathrm{Fix}(\Sigma) \neq 0$ for any isotropy subgroup $\Sigma \subsetneq \Gamma$, because either we have $\mathrm{Fix}(\Sigma) = \{\mathbf{0}\}$, but then the full isotropy subgroup must in fact be Γ because we have $\gamma\mathbf{0} = \mathbf{0}$ for any $\gamma \in \Gamma$, or we have some nonzero vector, $\boldsymbol{x}$, fixed by Σ and hence by equation (4.25) any scalar multiple of $\boldsymbol{x}$ is also fixed, and so we have $\dim \mathrm{Fix}(\Sigma) \geq 1$

4.1.4 Axial and maximal isotropy subgroups

An isotropy subgroup, Σ, satisfying

$$\dim \mathrm{Fix}(\Sigma) = 1 \tag{4.29}$$

is said to be **axial**

An isotropy subgroup, $\Sigma_x \subseteq \Gamma$, is said to be **maximal** if there is no isotropy subgroup $\Sigma_y \subset \Gamma$ satisfying

$$\Sigma_x \subsetneq \Sigma_y \subsetneq \Gamma, \tag{4.30}$$

in other words if it is not a subgroup of a bigger isotropy subgroup apart from the whole group Γ.

Lemma 4.2 *Let* $\mathrm{Fix}(\Gamma) = \{\mathbf{0}\}$, *and let* $\Sigma \subsetneq \Gamma$ *be a subgroup Then* Σ *is a maximal isotropy subgroup of* Γ *if and only if*

$$\dim \mathrm{Fix}(\Sigma) > 0, \tag{4.31}$$

$$\dim \mathrm{Fix}(\Delta) = 0 \quad \text{for every closed subgroup } \Delta \supsetneq \Sigma. \tag{4.32}$$

The proof of this lemma can be found in section 4.6.

Axial isotropy subgroups must be maximal. To see this, suppose that Σ is an axial isotropy subgroup that is not maximal. Since Σ is not maximal there must be an isotropy subgroup, Δ, satisfying $\Gamma \supsetneq \Delta \supsetneq \Sigma$, and so $\mathrm{Fix}(\Delta) \subsetneq \mathrm{Fix}(\Sigma)$ $\mathrm{Fix}(\Sigma)$ is one-dimensional because Σ is axial, and so it must be the space spanned by some nonzero vector, $\boldsymbol{x}$, since any scalar multiple of a vector, $\boldsymbol{x}$, in the fixed-point subspace will also lie in the fixed-point subspace Thus the only possibility is that $\mathrm{Fix}(\Delta) = \{\mathbf{0}\}$, which is a contradiction since $\Delta \neq \Gamma$

We can use the **trace formula** to compute $\dim \mathrm{Fix}(\Sigma)$ and so work out which isotropy subgroups are axial and/or maximal.

Theorem 4.3 (Trace formula) *Let* Γ *be a finite group, and let* $\Sigma \subset \Gamma$ *be a subgroup Then the dimension of the fixed-point subspace of* Σ *is given by*

$$\dim \mathrm{Fix}(\Sigma) = \frac{1}{|\Sigma|} \sum_{\sigma \in \Sigma} tr(\sigma), \tag{4.33}$$

where $tr(\sigma)$ *is the trace of the matrix representing* σ.

The proof of this theorem can be found in section 4.6 There is also a version of this theorem for compact Lie groups, which is given in Golubitsky, Stewart and Schaeffer (1988).

4.2 The equivariant branching lemma

The equivariant branching lemma is a very useful result which makes predictions about the symmetry of solutions at steady bifurcations, based on the symmetry of the bifurcation problem. It was proved by Vanderbauwhede (1980) and Cicogna (1981). We start by stating and proving a more general version of the lemma.

Theorem 4.4 (Generalised equivariant branching lemma) *Let Γ be a finite group or a compact Lie group acting on a real vector space, V, with* $\mathrm{Fix}(\Gamma) = \{\mathbf{0}\}$ *Let*

$$\frac{d\boldsymbol{x}}{dt} = f(\boldsymbol{x}, \mu) \tag{4.34}$$

be a Γ-equivariant bifurcation problem with $Df\big|_{(\mathbf{0},0)} = 0$ and $Df_\mu\big|_{(\mathbf{0},0)}\boldsymbol{v} \neq \mathbf{0}$ for nonzero $\boldsymbol{v} \in \mathrm{Fix}(\Sigma)$, where Σ is an isotropy subgroup of Γ. Then, if Σ satisfies

$$\dim \mathrm{Fix}(\Sigma) = 1, \tag{4.35}$$

there is a smooth solution branch $\boldsymbol{x} = s\boldsymbol{v}$, $\mu = \mu(s)$, to $f(\boldsymbol{x}, \mu) = \mathbf{0}$.

Proof Since fixed-point subspaces are flow-invariant we have $f(\boldsymbol{x}, \mu) \in \mathrm{Fix}(\Sigma)$ for all $\boldsymbol{x} \in \mathrm{Fix}(\Sigma)$. Since $\dim \mathrm{Fix}(\Sigma) = 1$ we also have $\boldsymbol{x} \parallel \boldsymbol{v}$ for all $\boldsymbol{x} \in \mathrm{Fix}(\Sigma)$ and so $\boldsymbol{x} = s\boldsymbol{v}$ and

$$f(s\boldsymbol{v}, \mu) = h(s, \mu)\boldsymbol{v}, \tag{4.36}$$

where $h(s, \mu)$ is a scalar function and s is a scalar that parameterises the solution branch. Now, since $\mathrm{Fix}(\Gamma) = \{\mathbf{0}\}$, there is a trivial solution $f(\mathbf{0}, \mu) = \mathbf{0}$. The vector $\boldsymbol{v}$ is nonzero, so the trivial solution corresponds to $s = 0$ and hence we must have $h(0, \mu) = 0$. Expanding $h(s, \mu)$ in equation (4.36) as a Taylor series in s gives

$$f(s\boldsymbol{v}, \mu) = h(0, \mu)\boldsymbol{v} + \left.\frac{\partial h}{\partial s}\right|_{s=0} s\boldsymbol{v} + \frac{1}{2}\left.\frac{\partial^2 h}{\partial s^2}\right|_{s=0} s^2\boldsymbol{v} + \tag{4.37}$$

$$= k(s, \mu)s\boldsymbol{v}, \tag{4.38}$$

where

$$k(s, \mu) = \left.\frac{\partial h}{\partial s}\right|_{s=0} + \frac{1}{2}s\left.\frac{\partial^2 h}{\partial s^2}\right|_{s=0} + \cdots \tag{4.39}$$

Now

$$k(0, 0)\boldsymbol{v} = Df\big|_{(\mathbf{0}\,0)}\boldsymbol{v} = \mathbf{0}, \tag{4.40}$$

$$\frac{\partial k}{\partial \mu}(0, 0)\boldsymbol{v} = Df_\mu\big|_{(\mathbf{0}\,0)}\boldsymbol{v} \neq \mathbf{0} \tag{4.41}$$

hold by assumption. Since $v \neq \mathbf{0}$ we have $k(0,0) = 0$ and $(\partial k/\partial \mu)(0,0) \neq 0$, so by the implicit function theorem (see Glendinning, 1994, if you are not familiar with this), for s and μ sufficiently small we can solve $k(s,\mu) = 0$ to find the unique solution for μ as a function of s. Since $\dim \mathrm{Fix}(\Sigma) = 1$, v is unique for a given Σ, up to scaling. Thus, given Σ, the solution branch is unique. The solution branch also has $x = sv \in \mathrm{Fix}(\Sigma)$, and so the isotropy subgroup of each solution x is Σ. □

The standard statement of the equivariant branching lemma characterises the unique smooth solution branch to $f(x,\mu) = \mathbf{0}$ in terms of the isotropy subgroup, Σ, rather than the form $(sv, \mu(s))$ of the solution branch, but clearly these two approaches are equivalent.

If we require that Γ act absolutely irreducibly and nontrivially on V, then it follows automatically that $\mathrm{Fix}(\Gamma) = \{\mathbf{0}\}$, and in fact the standard equivariant branching lemma is stated for absolutely irreducible actions, without explicit reference to $\mathrm{Fix}(\Gamma)$. The generalised version of the lemma, on the other hand, holds even for reducible actions as long as $\mathrm{Fix}(\Gamma) = \{\mathbf{0}\}$.

It can be shown that for a generic steady-state Γ-equivariant bifurcation problem the action of Γ on the centre eigenspace is absolutely irreducible. The proof is rather involved: a sketch is given in Golubitsky, Stewart and Schaeffer (1988). A very brief outline of the ideas behind it is the following. If the representation of Γ were reducible and orthogonal, then in suitable coordinates we could write

$$Df|_{x=\mathbf{0}} = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}, \tag{4.42}$$

where A acts on a Γ-invariant subspace, U, and B acts on its Γ-invariant complement, $U^{\perp}$, and $x = \mathbf{0}$ is the fixed point that loses stability at the bifurcation. As we are on the centre eigenspace, the real parts of the eigenvalues of A and B are zero at the bifurcation point $\mu = 0$. Now add a small perturbation to give

$$Df|_{x=\mathbf{0}} = \begin{pmatrix} A & 0 \\ 0 & B + \epsilon I \end{pmatrix}, \tag{4.43}$$

where $0 < \epsilon \ll 1$. The eigenvalues of $Df|U$ are unchanged, but on the subspace $U^{\perp}$ they are no longer zero at $\mu = 0$. A centre manifold reduction produces a new centre manifold, tangent to U at $x = \mathbf{0}$, which is of lower dimension than the original, $U \oplus U^{\perp}$, so the unperturbed bifurcation problem cannot be generic. A generic bifurcation must therefore be governed by an irreducible representation. Now if Γ acts irreducibly, but not absolutely irreducibly then there must be matrices other than multiples of the identity that commute with the group elements. It turns out that the only possibilities are for the matrices to be of **complex type** or

quaternionic type. The space of complex-type matrices is isomorphic to the complex numbers, while the space of quaternionic-type matrices is isomorphic to the quaternions Since Df commutes with Γ it must also take one of these forms or be a multiple of the identity. An example of a complex-type matrix is

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \quad a, b \in \mathbb{R}, \tag{4.44}$$

which is the form taken by elements of $SO(2)$ in Example 3 26 If $Df|_{x=0}$ is of complex type, then its eigenvalues come in complex conjugate pairs, $\sigma(\mu) \pm i\omega(\mu)$. If the bifurcation at $\mu = 0$ is to be steady then we must have $\sigma(0) = \omega(0) = 0$ In the example given in (4 44), the eigenvalues are $a \pm ib$, with $a = b = 0$ at $\mu = 0$. It is then possible to add a small perturbation of complex type to get new eigenvalues $\sigma(\mu) \pm i(\omega(\mu) + \epsilon)$, where $0 < \epsilon \ll 1$. In the example, we could set

$$Df|_{x=0} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} + \begin{pmatrix} 0 & -\epsilon \\ \epsilon & 0 \end{pmatrix}, \tag{4.45}$$

and the new eigenvalues would be $a \pm i(b + \epsilon)$ The bifurcation still occurs at $\mu = 0$, where the eigenvalues are $\pm i\epsilon$, but it is no longer steady In fact there is no longer a steady bifurcation close to $\mu = 0$ at all Thus if the bifurcation is to be generic, Df cannot be of complex type at a steady bifurcation Similarly it is possible to rule out Df being of quaternionic type, but since such representations do not arise naturally in real applications, we won't go into the details here. The only remaining possibility is that Df must be a multiple of the identity and so generically the group acts absolutely irreducibly. This explanation has deliberately swept lots of messy details under the carpet, so look in Golubitsky, Stewart and Schaeffer (1988) if you want to know more

If Γ acts absolutely irreducibly, then by definition the only matrices commuting with all $\gamma \in \Gamma$ are scalar multiples of the identity, so if we differentiate the equivariance condition

$$f(\gamma x, \mu) = \gamma f(x, \mu), \quad \forall \gamma \in \Gamma \tag{4.46}$$

to obtain

$$Df\big|_{(0,\mu)}\gamma = \gamma Df\big|_{(0,\mu)}, \quad \forall \gamma \in \Gamma \tag{4.47}$$

we see that we must have $Df\big|_{(0,\mu)} = c(\mu)I$ for some scalar $c(\mu)$ This excludes the possibility of a Hopf bifurcation, where we would have a pair of purely imaginary eigenvalues, $\pm i\lambda$, $\lambda \in \mathbb{R}$ at the bifurcation point. (Hopf bifurcations can be associated with nonabsolutely irreducible representations of complex type, as we shall see in Section 4 4.) Similarly it excludes the possibility that

equation (4 1) is Hamiltonian. To see this, consider the two-dimensional Hamiltonian equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x, \mu) \equiv \begin{pmatrix} \frac{\partial H}{\partial x_2} \\ -\frac{\partial H}{\partial x_1} \end{pmatrix}, \tag{4.48}$$

where $x = (x_1, x_2)$ and $H \equiv H(x_1, x_2)$. The stability of a fixed point, $x = x_0$, is governed by the eigenvalues of the Jacobian

$$Df|_{x_0} \equiv \left. \begin{pmatrix} \frac{\partial^2 H}{\partial x_1 \partial x_2} & \frac{\partial^2 H}{\partial x_2^2} \\ -\frac{\partial^2 H}{\partial x_1^2} & -\frac{\partial^2 H}{\partial x_1 \partial x_2} \end{pmatrix} \right|_{(x_0\ \mu)}, \tag{4 49}$$

which are

$$\lambda_{\pm} = \pm \sqrt{\left(\left. \frac{\partial^2 H}{\partial x_1 \partial x_2} \right|_{(x_0, \mu)} \right)^2 - \left. \frac{\partial^2 H}{\partial x_1^2} \right|_{(x_0, \mu)} \left. \frac{\partial^2 H}{\partial x_2^2} \right|_{(x_0\ \mu)}}. \tag{4 50}$$

The eigenvalues at any fixed point come in pairs $\pm\lambda$ or $\pm i\lambda$, $\lambda \in \mathbb{R}$, so the Jacobian cannot be rewritten in the form $Df = c(\mu)I$, and Γ cannot act irreducibly. The argument can be extended easily to higher dimensional Hamiltonian systems

In the generalised statement of the equivariant branching lemma the requirement $Df_{\mu}|_{(\mathbf{0}\ 0)} v \neq \mathbf{0}$ gives a different non-degeneracy condition for each choice of v and hence for each solution branch. Now we have $Df_{\mu}|_{(\mathbf{0},0)} v = c'(0)v$, so that requirement is replaced by

$$c'(0) \neq 0, \tag{4 51}$$

which holds simultaneously for all isotropy subgroups, Σ, of Γ.

We can now state the equivariant branching lemma in its standard form.

Theorem 4.5 (Equivariant branching lemma) *Let Γ be a finite group or compact Lie group acting absolutely irreducibly on a real vector space, V, and let*

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x, \mu) \tag{4.52}$$

be a Γ-equivariant bifurcation problem with $f(\mathbf{0}, 0) = \mathbf{0}$ and $Df|_{(\mathbf{0},0)} = 0$ that satisfies equation (4 51). If Σ is an isotropy subgroup of Γ, satisfying

$$\dim \mathrm{Fix}(\Sigma) = 1, \tag{4 53}$$

then there exists a unique smooth solution branch to $f(x, \mu) = \mathbf{0}$ such that the isotropy subgroup of each solution is Σ

What is really going on here? Well, if the flow-invariant subspace $\mathrm{Fix}(\Sigma)$ is one-dimensional, then the bifurcation problem is one-dimensional there. Since $\mathrm{Fix}(\Gamma) = \{\mathbf{0}\}$, there is an equilibrium solution $f(\mathbf{0}, \mu) = \mathbf{0}$. The generic one-dimensional steady-state bifurcation with a fixed point at $x = \mathbf{0}$ is either transcritical (in the absence of symmetry) or pitchfork (if there is symmetry under $x \to -x$). So the equivariant branching lemma says that by restricting to $\mathrm{Fix}(\Gamma)$ you will end up with a one-dimensional transcritical or pitchfork bifurcation.

There may also be solution branches that bifurcate from the origin with isotropy subgroup, Σ, such that $\dim \mathrm{Fix}(\Sigma) > 1$. The equivariant branching lemma says nothing about them however, and in general they will have to be found directly from equation (4.1). In fact Cicogna (1981) proved that a solution branch must exist if $\dim \mathrm{Fix}(\Sigma)$ is odd; in this case the solutions will have isotropy subgroup, Σ, if Σ is maximal.

All branches that bifurcate from the origin are known as **primary** branches, but only those that are predicted by the equivariant branching lemma are **axial** branches. Sometimes nonaxial primary branches are forced to exist for all combinations of coefficients in the normal form equations, and sometimes they are found only in certain regions of coefficient space.

Equivariance can also give us some information about the stability of the bifurcating solution branches. It can be shown (see Golubitsky, Stewart and Schaeffer, 1988, for the proof) that, under the hypotheses of the equivariant branching lemma, if a guaranteed solution branch is transcritical, or if certain conditions on the quadratic part of $f(x, \mu)$ are satisfied in $\mathrm{Fix}(\Sigma)$, then the branch is unstable. In this book we will investigate stability using the governing equations directly.

The next thing to do is to see how the equivariant branching lemma applies in practice, so we will move on and consider some worked examples.

4.3 Bifurcations in a box

In general, when we think of patterns, we tend to imagine more or less periodic structures with many repeats of the pattern in some large domain, such as you might see in animal coat markings or convection experiments. In those cases, the appropriate bifurcation problem usually takes place on a lattice, and the symmetry group is related to that of the lattice itself. Those are pretty complex problems to start with, though, so we defer all thoughts of lattices until the next chapter. It is perfectly possible for solutions with less symmetry to appear in small domains: this is the kind of thing that might happen in a convection or reaction-diffusion experiment in a small box, such as the one shown

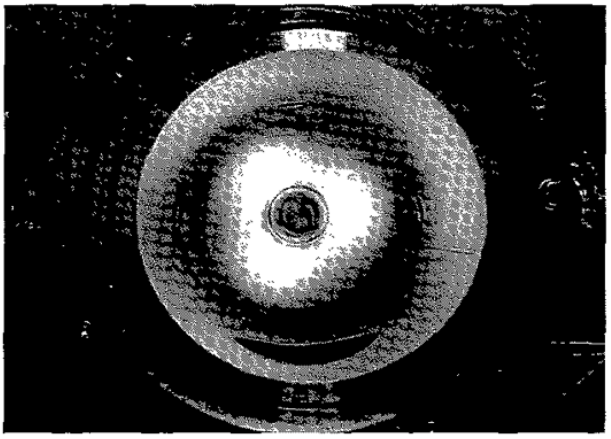


Fig. 4.2 A pattern with approximate D_3 symmetry in a toroidal plasma column experiment. The bright blobs in the centre show spatial modulations of the plasma density and temperature. Photograph courtesy of Dana Mackey and Angus McCarter, Plasma Research Laboratory, Dublin City University; ©Angus McCarter, 2003

in Figure 4.2. The symmetry groups here are much smaller and more manageable, so we will choose this type of problem for our first application of the equivariant branching lemma.

4.3.1 Steady bifurcation in a square box

Imagine a pattern-forming experiment in a small square box. The box has symmetry group D_4, the symmetry group of a square, and so the bifurcation problem will be D_4-equivariant as long as no additional symmetries are generated by the boundary conditions acting at the edges of the box. The possibility of such **hidden** symmetries will be discussed later, in Section 6.6. Taking D_4 as the symmetry group of the bifurcation problem, and assuming that the bifurcation is steady, we can go ahead and work out the isotropy subgroups, find out which ones have a one-dimensional fixed-point subspace, and hence discover the symmetries of the primary solution branches.

There are many different representations of the symmetry group, and the choice of representation will affect the isotropy subgroups and the dimensions of the fixed-point subspaces, so we will need to decide which representation to take before we start. Since a generic steady-state symmetric bifurcation is governed by an absolutely irreducible real representation, we shall identify all such representations of D_4 and see what the equivariant branching lemma tells us in each case. In any given real-life experiment, only one representation will be relevant,

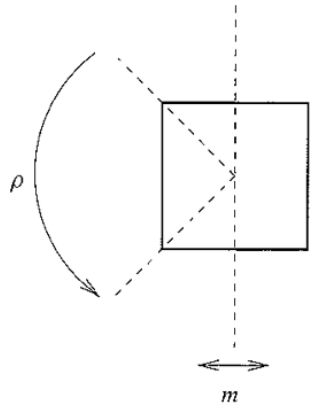


Fig. 4.3 The generators of $\mathcal{D}_4$, the symmetry group of a square, are m, a reflection, and ρ, a rotation through $\pi/2$

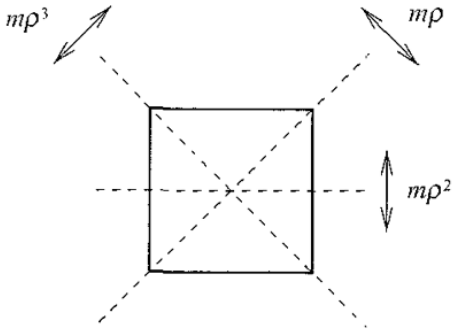


Fig. 4.4 The reflections $m\rho$, $m\rho^2$ and $m\rho^3$ of $\mathcal{D}_4$

but if we find out what happens for each of the irreps we will have a better chance of identifying the right representation for our particular application by comparing our results with the range of results predicted by theory. Of course, we may be unlucky: our experiment may have 'chosen' a very strange or unusual representation of the symmetry group, and we may have to do a bit of detective work to figure out what it is, but in general nature is relatively helpful in these matters and the absolutely irreducible representations crop up quite often.

The group $\mathcal{D}_4$ is generated by a reflection m in the y axis and ρ, a rotation by $\pi/2$ as shown in Figure 4.3. The group elements are e, ρ, ρ^2, ρ^3, m, $m\rho$, $m\rho^2$ and $m\rho^3$. The elements ρ^2 and ρ^3 are rotations through π and $3\pi/2$ respectively, while the elements $m\rho$, $m\rho^2$ and $m\rho^3$ are reflections, illustrated in Figure 4.4.

Table 4.1 *The one-dimensional irreps of D_4*

Irrep	e	ρ	ρ^2	ρ^3	m	$m\rho$	$m\rho^2$	$m\rho^3$
R_1	1	1	1	1	1	1	1	1
R_2	1	1	1	1	−1	−1	−1	−1
R_3	1	−1	1	−1	1	−1	1	−1
R_4	1	−1	1	−1	−1	1	−1	1

The natural representation of the group is given by the set of matrices

$$M_e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad M_\rho = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$
$$M_{\rho^2} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad M_{\rho^3} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$
$$M_m = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad M_{m\rho} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$
$$M_{m\rho^2} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad M_{m\rho^3} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \tag{4.54}$$

which transform the Cartesian plane in the same way as the group elements transform the square. This is the representation we would naturally pick if asked to associate some matrices with the group elements. We would use coordinate geometry to figure out the matrices, M_ρ, for a rotation of the plane by $\pi/2$, and M_m, a reflection in the y axis, and then multiply them together in the appropriate combinations to find the matrices corresponding to all the other group elements. This gives us a representation of D_4. It is a two-dimensional absolutely irreducible real representation. However, this is not the only possibility: there are also one-dimensional absolutely irreducible real representations, as given in Table 4.1.

You can check that the one-dimensional irreps are indeed representations, by looking to see whether they obey the group structure. For example, since $m\rho = m \cdot \rho$ any representation should satisfy $M_{m\rho} = M_m M_\rho$. Considering irrep R_2, we see that $M_{m\rho} = -1 = -1 \times 1 = M_m M_\rho$, and so these matrices behave according to the group structure. In fact you can check that any such equation is satisfied by the matrices of each of the one-dimensional irreps given in Table 4.1. As the representations are one-dimensional they automatically satisfy the condition for absolute irreducibility, that the only matrices to commute with all the matrices of the representation are scalar multiples of the identity.

This example suggests a method of finding the absolutely irreducible representations of a finite group. First identify the generators γ_i, for $i = 1, \ldots, n$, of the

group. If the order of the generator γ_i is p, then $\gamma_i^p = e$. Hence we must have $M_{\gamma_i}^p = I$, and so the matrix M_{γ_i} must be a p^{th} root of I. Once you have identified matrices to represent the generators, you can work out all the remaining matrices using matrix multiplication. The combination of all the possible matrices for all the generators will give you all the irreps, although in general you will find that some of the choices of matrices do not give representations. In addition you might actually find more representations than you need because some will be equivalent to combinations of the others, and so not irreducible, and some will be conjugate to others and so equivalent. Of course, some might be complex, and so not relevant to a steady bifurcation unless they are equivalent to a real absolutely irreducible representation under a change of basis. In Section 4.3.2 we will see how to identify the inequivalent irreps, but for now we shall take a look at how the method works for the one-dimensional irreps of $\mathcal{D}_4$. The generators are m, of order two, and ρ, of order four. Thus for the matrices of the irreps we must have

$$M_m^2 = 1, \quad M_\rho^4 = 1, \tag{4.55}$$

and so $M_m = \pm 1$ and $M_\rho = \pm 1, \pm i$. Table 4.1 was constructed by taking all four combinations of $M_m = \pm 1$ and $M_\rho = \pm 1$. Why do we not use $M_\rho = \pm i$? Well, in that case we would have $M_{m\rho} = \pm i$ and hence $M_{m\rho}^2 = -1$, but we know that $(m\rho)^2 = e$ since $m\rho$ is a reflection, and so we must have $M_{m\rho}^2 = 1$. In other words, the choices $M_\rho = \pm i$ do not lead to representations of the group, and so we discard them. We are left with only real one-dimensional irreps in this example, but that will not always be the case: for example, the group $\mathcal{C}_4$, with generator a satisfying $a^4 = e$, has an irrep where $M_a = i$.

Each irrep leads to a corresponding bifurcation problem with normal form

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = f(\boldsymbol{x}, \mu), \tag{4.56}$$

where $f(\boldsymbol{x}, \mu)$ satisfies

$$Mf(\boldsymbol{x}, \mu) = f(M\boldsymbol{x}, \mu), \tag{4.57}$$

for each matrix M in the representation. If M is an $n \times n$ matrix, we have $\boldsymbol{x} \in \mathbb{R}^n$. Each component of $\boldsymbol{x}$ is the amplitude of an eigenmode represented by the corresponding basis vector. The eigenmode behaves under the symmetries of the group as the basis vector does under the matrices of the representation. The equivariance condition (4.57) determines the form of $f(\boldsymbol{x}, \mu)$, but there will be free parameters that are not fixed by the symmetries. For any given experiment or theoretical application, the values of these parameters can be determined from measurements or from the governing equations, as we will see in Chapter 7.

Under irrep R_1, we have $x \in \mathbb{R}$ and $\mathrm{Fix}(D_4) = \mathbb{R}$. The representation acts irreducibly, but its fixed-point subspace is the whole space, $\mathbb{R}$, rather than $\{0\}$. The conditions for the equivariant branching lemma to hold are not met in this case, but since the group action is trivial it is easy to work out what happens. The same will hold for the identity irrep of any group.

Since all the matrices of the representation are $+1$ in this case, the equivariance condition (4.57) reduces to

$$1 \cdot f(x, \mu) = f(1 \cdot x, \mu), \tag{4.58}$$

and so does not restrict the form of $f(x, \mu)$. If there is a fixed point at $x = 0$, $\mu = 0$, that undergoes a steady-state bifurcation at $\mu = 0$, then we must have $f(0, 0) = 0$ and $\mathrm{d}f/\mathrm{d}x(0, 0) = 0$. For small x and μ, $f(x, \mu)$ must therefore take the form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mu + ax^2 + bx\mu + c\mu^2 + \quad , \tag{4.59}$$

where a, b and c are constants independent of μ. Generically we expect $a \neq 0$ at $\mu = 0$. Close to the bifurcation, where μ is small, the leading-order balance is then between μ and x^2, and so we scale $x \sim \mu^{\frac{1}{2}}$, assuming that a is $O(1)$. Then the terms $x\mu$, μ^2 and so on only appear at higher order in the expansion, and we can neglect them at leading order, giving the equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mu + ax^2 + \tag{4.60}$$

From now on we shall use similar arguments implicitly to write down leading-order amplitude equations where all the coefficients are assumed independent of μ unless otherwise stated.

Equation (4.60) describes a saddle-node bifurcation. Under the identity representation every point x has all the symmetries of the group, so the branch of stationary solutions must have full D_4 symmetry. In other words the isotropy subgroup of solutions on this branch is D_4.

For the irreps R_2, R_3 and R_4 we have $x \in \mathbb{R}$. Some of the matrices in each irrep are $+1$, which when substituted into the equivariance condition (4.57) give

$$f(x, \mu) = f(x, \mu), \tag{4.61}$$

and so impose no restriction on the form of $f(x, \mu)$. The remaining matrices in each irrep are -1. In these cases the equivariance condition (4.57) becomes

$$-f(x, \mu) = f(-x, \mu) \tag{4.62}$$

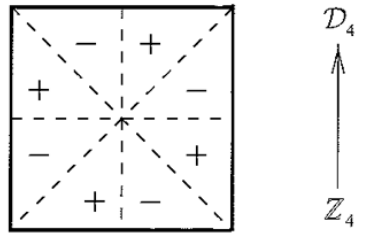


Fig 4.5 (a) Solution eigenmode for the irrep R_2 of $\mathcal{D}_4$ and (b) isotropy lattice for the irrep R_2 of $\mathcal{D}_4$, where inclusion is shown by an arrow

This tells us that the function $f(x, \mu)$ is odd in x, and so expanding in powers of x gives us the amplitude equation

$$\frac{dx}{dt} = \mu x + ax^3 + \cdots \tag{4.63}$$

This describes a pitchfork bifurcation, which will be subcritical if $a > 0$ and supercritical if $a < 0$

In these three cases, the basis vector is just 1, which is unchanged under multiplication by all the matrices $+1$ of the irrep and sent to -1 under all the matrices -1 of the irrep. The corresponding solution eigenmode must therefore remain unchanged under all the symmetries represented by $+1$ in the irrep, and must be sent to its negative under all the symmetries represented by -1. For example, the solution eigenmode for R_2 must be unchanged under the group elements e, ρ, ρ^2 and ρ^3, because they correspond to matrices $+1$ in the representation, and it must be transformed into its negative under the group elements m, $m\rho$, $m\rho^2$ and $m\rho^3$, because they correspond to matrices -1. An example is shown in Figure 4.5a: its isotropy subgroup, $\{e, \rho, \rho^2, \rho^3\} \cong \mathbb{Z}_4$, consists of all the elements that leave the solution unchanged, namely those that correspond to matrices $+1$. The isotropy subgroup has a one-dimensional fixed-point subspace since there is only one degree of freedom in the solution: the size of the amplitude, x, of the eigenmode. The lattice of isotropy subgroups for this action of D_4 on $\mathbb{R}$ is shown in Figure 4.5b. The origin, $x = 0$, has isotropy subgroup D_4, and all other points $x \in \mathbb{R}$ have isotropy subgroup $\mathbb{Z}_4$.

Now D_4 is finite. We also have $\mathrm{Fix}(D_4) = \{0\}$ under R_2, and we assume the form of the bifurcation problem specified in the statement of the equivariant branching lemma. So a branch of solutions with isotropy subgroup $\mathbb{Z}_4 = \{e, \rho, \rho^2, \rho^3\}$, bifurcating from the origin at $\mu = 0$, is guaranteed by the equivariant branching lemma.

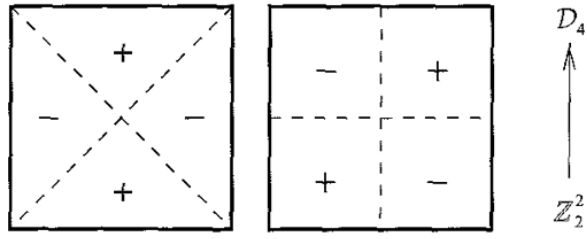


Fig. 4.6. Solution eigenmodes for the irreps (a) R_3 and (b) R_4 of $\mathcal{D}_4$, and (c) the isotropy lattice for the irreps R_3 and R_4 of $\mathcal{D}_4$, where inclusion is shown by an arrow

Using the same procedure it is now easy to work out that the eigenmode for R_3 has isotropy subgroup $\{e, \rho^2, m, m\rho^2\} \cong \mathbb{Z}_2^2$, and the one for R_4 has isotropy subgroup $\{e, \rho^2, m\rho, m\rho^3\} \cong \mathbb{Z}_2^2$. Both have a one-dimensional fixed-point subspace, and so branches of solutions with these symmetries are guaranteed by the equivariant branching lemma. Suitable eigenmodes are shown in Figures 4.6a and 4.6b. The isotropy lattice is the same for these two actions, and is shown in Figure 4.6c. The origin, $x = 0$, has isotropy subgroup $\mathcal{D}_4$, and all other points $x \in \mathbb{R}$ have isotropy subgroup $\mathbb{Z}_2^2$.

We can check that the trace formula holds. In all four one-dimensional cases, the isotropy subgroup, Σ, consists of all those elements with character +1, so we have

$$\dim \mathrm{Fix}(\Sigma) = \frac{1}{|\Sigma|} \sum_{\sigma \in \Sigma} tr(\sigma) = \frac{1}{|\Sigma|} \sum_{\sigma \in \Sigma} 1 = 1, \tag{4.64}$$

confirming that the fixed-point subspace is one-dimensional.

Finally, consider the natural representation of $\mathcal{D}_4$ that we introduced at the beginning of this section. We shall call it R_5. The vector of mode amplitudes, $\boldsymbol{x}$, is two-dimensional in this case because we are dealing with a two-dimensional irrep. Thus the amplitude equations take the form

$$\begin{pmatrix} \mathrm{d}x_1/\mathrm{d}t \\ \mathrm{d}x_2/\mathrm{d}t \end{pmatrix} = \begin{pmatrix} f_1(x_1, x_2, \mu) \\ f_2(x_1, x_2, \mu) \end{pmatrix}. \tag{4.65}$$

The form of $f_1(x_1, x_2, \mu)$ and $f_2(x_1, x_2, \mu)$ can be deduced by expanding them in powers of x_1 and x_2, and applying the equivariance condition (4.57) using the matrices corresponding to the generators of the group. Applying the matrix

$$M_m = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{4.66}$$

for the reflection m shows that

$$\begin{pmatrix} -f_1(x_1, x_2, \mu) \\ f_2(x_1, x_2, \mu) \end{pmatrix} = \begin{pmatrix} f_1(-x_1, x_2, \mu) \\ f_2(-x_1, x_2, \mu) \end{pmatrix} \tag{4.67}$$

must hold, so f_1 must be odd and f_2 must be even in x_1. Applying the rotation matrix

$$M_\rho = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{4.68}$$

gives

$$\begin{pmatrix} -f_2(x_1, x_2, \mu) \\ f_1(x_1, x_2, \mu) \end{pmatrix} = \begin{pmatrix} f_1(-x_2, x_1, \mu) \\ f_2(-x_2, x_1, \mu) \end{pmatrix}. \tag{4.69}$$

From the second row we have

$$f_1(x_1, x_2, \mu) = f_2(-x_2, x_1, \mu) = f_2(x_2, x_1, \mu) = f_1(x_1, -x_2, \mu), \tag{4.70}$$

since f_2 is even in the first argument. Hence f_1 is even in the second argument. Thus up to cubic order, f_1 must take the form

$$f_1(x_1, x_2, \mu) = \mu x_1 - a_1 x_1^3 - a_2 x_2^2 x_1, \tag{4.71}$$

where a_1 and a_2 are real constants, and μ is the real bifurcation parameter. Now using equation (4.69) we can deduce the normal form to be

$$\begin{pmatrix} \mathrm{d}x_1/\mathrm{d}t \\ \mathrm{d}x_2/\mathrm{d}t \end{pmatrix} = \mu \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - a_1 \begin{pmatrix} x_1^3 \\ x_2^3 \end{pmatrix} - a_2 \begin{pmatrix} x_2^2 x_1 \\ x_1^2 x_2 \end{pmatrix}. \tag{4.72}$$

If we now make the substitution $\widehat{x_i} = |a_1|^{\frac{1}{2}} x_i$, $\widehat{a_2} = a_2/|a_1|$ and immediately drop the hats, the normal form is transformed to

$$\begin{pmatrix} \mathrm{d}x_1/\mathrm{d}t \\ \mathrm{d}x_2/\mathrm{d}t \end{pmatrix} = \mu \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mp \begin{pmatrix} x_1^3 \\ x_2^3 \end{pmatrix} - a_2 \begin{pmatrix} x_2^2 x_1 \\ x_1^2 x_2 \end{pmatrix}, \tag{4.73}$$

where the minus sign holds for $a_1 > 0$ and the plus sign for $a_1 < 0$. Another way to describe this procedure is to say that we are setting $a_1 = \pm 1$ by rescaling. It is helpful because it reduces the number of free parameters by one. Very often in a particular application there will be features that fix the sign of a_1 too. Typically we choose $a_1 = 1$ because then each amplitude x_i independently has a nonzero stationary solution in $\mu > 0$, which is physically reasonable for most applications. For the rest of this section, though, we will not fix a_1, and we will work with equations (4.72).

There are two eigenmodes, whose amplitudes are measured by x_1 and x_2, that behave under the symmetries of the group as the basis vectors $(1, 0)$ and

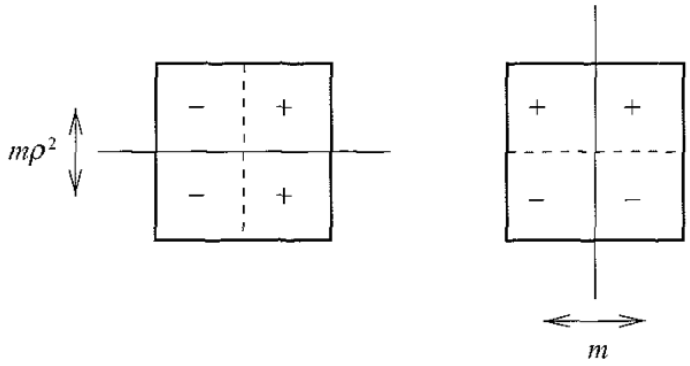


Fig 4 7. Eigenmodes foi the natural iepresentation of $\mathcal{D}_4$

$(0, 1)$ do under the matrices of the natural representation They are shown in Figure 4 7 and their amplitudes are solutions of the normal form equation (4 72). The isotropy subgroup of the lefthand solution is given by $\mathbb{Z}_2 = \{e, m\rho^2\}$. It has a one-dimensional fixed-point subspace, and so is guaranteed by the equivariant branching lemma, since it corresponds to a solution $(x_1, 0)$ where only the amplitude, $x_1 = \pm\sqrt{\mu/a_1}$, can be varied (by varying μ). The righthand solution has isotropy subgroup $\{e, m\}$ and corresponds to a solution $(0, x_2)$ It has the same existence and stability properties as the one on the left because it is on the same group orbit: you can transform the lefthand picture into the one on the right by rotating through $\pi/2$. Equivalently, the isotropy subgroup $\{e, m\rho^2\}$ of the lefthand solution is conjugate to that of the one on the right since $\rho m\rho^{-1} = m\rho^2$ and $\rho e\rho^{-1} = e$

There are also solutions $x_1 = x_2$ and $x_1 = -x_2$, with $x_1^2 = \mu/(a_1 + a_2)$, which are shown in Figure 4.8 and have isotropy subgroups $\{e, m\rho\}$ and $\{e, m\rho^3\}$ respectively The fixed-point subspace is one-dimensional in each case, since both amplitudes, x_1 and x_2, vary together as μ is varied These two solutions are on the same group orbit (again rotate through $\pi/2$) and so are equivalent. Again the two isotropy subgroups are conjugate since $\rho(m\rho)\rho^{-1} = m\rho^3$.

Each of the isotropy subgroups $\{e, m\}$, $\{e, m\rho\}$, $\{e, m\rho^2\}$ and $\{e, m\rho^3\}$ fixes a line in the plane (the vertical centre line, bottom left to top right diagonal, horizontal centre line, and top left to bottom right diagonal respectively). These lines are their fixed-point subspaces and so it is clear that they are one-dimensional Here the connection is obvious because we are dealing with a two-dimensional irrep acting on a plane For the one-dimensional irreps it may not appear quite so obvious because we cannot point to a fixed line in a plane, but we can deduce that the isotropy subgroups have a one-dimensional fixed-point subspace because the symmetries do not fix the size of x

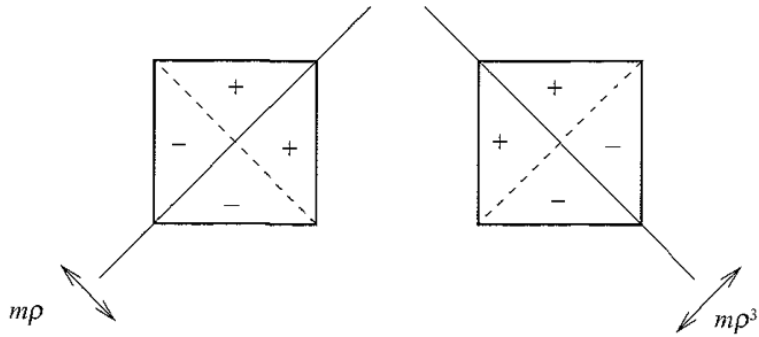


Fig. 4.8. Diagonal modes that have isotropy subgroups with one-dimensional fixed-point subspace in the natural representation of $\mathcal{D}_4$

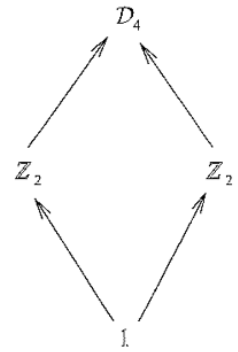


Fig. 4.9. Isotropy lattice for the natural representation of $\mathcal{D}_4$. Inclusion is shown by an arrow

The isotropy lattice for the natural representation of $\mathcal{D}_4$ is shown in Figure 4.9. The origin has isotropy subgroup $\mathcal{D}_4$, all points on an axis of reflection, other than the origin, have isotropy subgroup $\mathbb{Z}_2$, and all other points have isotropy subgroup $\mathbb{1}$ (consisting solely of the identity).

Again we can check that the trace formula holds. The trace of each matrix representing a reflection is zero, while the trace of the identity matrix is two. There are two elements, the identity and a reflection, in each isotropy subgroup Σ and so we have

$$\dim \mathrm{Fix}(\Sigma) = \frac{1}{|\Sigma|}\sum_{\sigma\in\Sigma} tr(\sigma) = \frac{1}{2}(2+0) = 1, \tag{4.74}$$

showing that the fixed-point subspace must be one-dimensional.

The stability of the solution $(x_1, 0)$, with $x_1^2 = \mu/a_1$, is governed by the eigenvalues of Df, evaluated at the solution. The Jacobian matrix, Df, can be calculated

from the normal form equation (4.72) and is given by

$$Df = \begin{pmatrix} \mu - 3a_1x_1^2 - a_2x_2^2 & -2a_2x_1x_2 \\ -2a_2x_1x_2 & \mu - 3a_1x_2^2 - a_2x_1^2 \end{pmatrix} \tag{4.75}$$

Evaluating this at $x_1^2 = \mu/a_1$, $x_2 = 0$, gives the diagonal matrix

$$Df = \begin{pmatrix} -2\mu & 0 \\ 0 & \mu(1 - a_2/a_1) \end{pmatrix}, \tag{4.76}$$

so the solution is stable to perturbations in x_1 if $\mu > 0$ and to perturbations in x_2 if $a_1 < a_2$ (since we must have $\mu/a_1 > 0$ for the solution to exist at all). The matrix is diagonal because we are using coordinates that correspond to the isotypic components with respect to $\Sigma_\lambda = \{e, m\rho^2\}$, the isotropy subgroup of the solution, which are

$$\begin{pmatrix} a \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ b \end{pmatrix}, \quad a, b \in \mathbb{R} \tag{4.77}$$

The other solution on the group orbit $(0, x_2)$, with $x_2^2 = \mu/a_1$, of course has the same stability properties, but now the nonzero entries in the Jacobian are swapped so that the eigenvalue -2μ corresponds to perturbations in the x_2 direction and $\mu(1 - a_2/a_1)$ to perturbations in the x_1 direction

If we evaluate the Jacobian at the solution $x_1 = x_2$, with $x_1^2 = \mu/(a_1 + a_2)$, we get

$$Df = \begin{pmatrix} -2a_1\mu/(a_1 + a_2) & -2a_2\mu/(a_1 + a_2) \\ -2a_2\mu/(a_1 + a_2) & -2a_1\mu/(a_1 + a_2) \end{pmatrix}, \tag{4.78}$$

which is not diagonal This is because the isotypic components with respect to the isotropy subgroup in this case, $\Sigma_\lambda = \{e, m\rho\}$, are different, namely

$$\begin{pmatrix} a \\ a \end{pmatrix} \text{ and } \begin{pmatrix} b \\ -b \end{pmatrix}, \quad a, b \in \mathbb{R}. \tag{4.79}$$

If we took coordinates along these directions, the matrix would be diagonal In other words, these should be eigenvectors of the Jacobian given in equation (4.78) It is easy to check that they are, with eigenvalues -2μ and $-2\mu(a_1 - a_2)/(a_1 + a_2)$ respectively, so the solution is stable to perturbations in the (1, 1) direction if $\mu > 0$, and to perturbations in the $(1, -1)$ direction if $a_1 > a_2$. Again, the other solution on the group orbit has the same stability properties, but with the roles of the eigenvectors swapped

Completing the stability analysis, the linearisation around the trivial solution, $x_1 = x_2 = 0$, shows that the solution is unstable for $\mu > 0$ and stable for $\mu < 0$

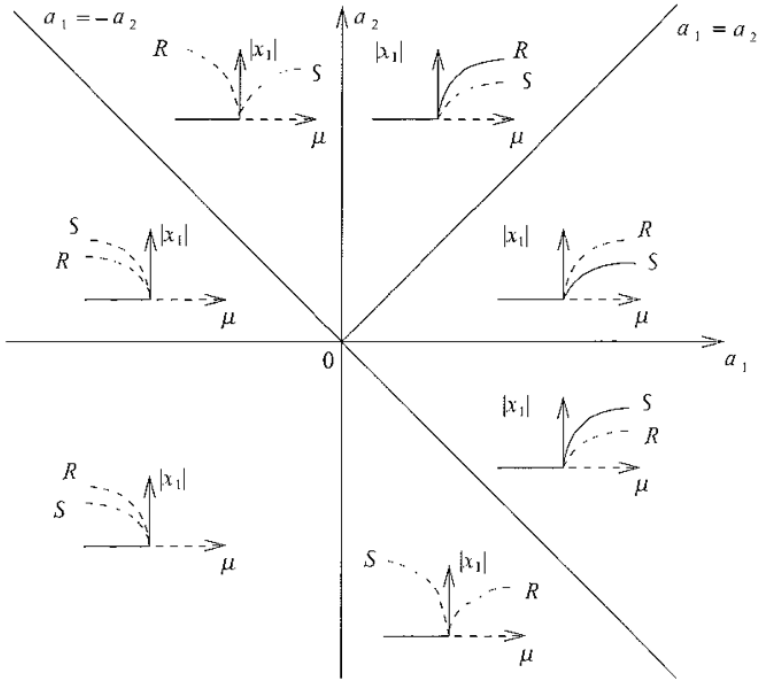


Fig. 4.10 Bifurcation diagrams for the bifurcation in a square box for various values of the coefficients a_1 and a_2 in equation (4.72). Solid lines represent stable solutions and dashed lines unstable solutions. Branches of solutions of type $(x, 0)$ and (x, x) are labelled R and S respectively.

The bifurcation diagrams for the various regions of (a_1, a_2) space are shown in Figure 4.10, where R denotes a solution of the first type, such as $(x, 0)$, and S denotes a solution of the second type, such as (x, x).

We have now catalogued all the solutions, characterised by their symmetries, that are guaranteed to appear as primary branches at a steady bifurcation with $\mathcal{D}_4$ symmetry under one or other of the absolutely irreducible actions of the group. (It turns out that there are no absolutely irreducible real representations other than R_1 to R_5, as we shall see in the next section.) For any given bifurcation problem, only one representation of the group will apply, and only the corresponding solution branches will be present. For example, Figure 4.11 shows a schematic representation of Bénard–Marangoni convection patterns that were observed by Ondarçuhu *et al.* (1993) in a square vessel. As the temperature of the bottom plate was increased, the D_4-symmetric pattern in Figure 4.11a lost stability to one or other of the patterns in Figures 4.11b or c. The new patterns are symmetric under reflection in both diagonals, so the bifurcation is governed by irrep R_4.

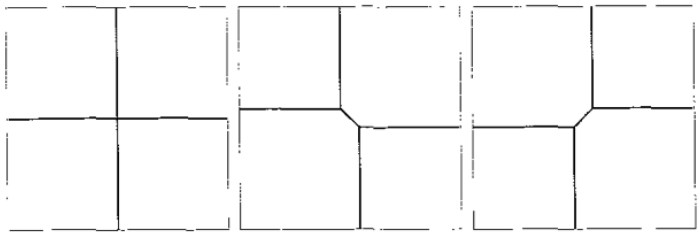


Fig 4 11. Ondarçuhu *et al* (1993) observed that a Bénard–Marangoni convection pattern of the form shown in (a) lost stability to one or other of the patterns (b) and (c) as the heating was increased The narrow lines show the edges of the convection vessel, while the thick lines show narrow regions where the fluid is moving downwards.

4.3.2 Finding the irreps

In order to discover all the possible primary branches for a generic steady Γ-equivariant bifurcation problem we first need to identify all the absolutely irreducible representations of Γ As we saw above, we can try to identify the matrices M_{γ_i} corresponding to generators γ_i of a finite group by solving

$$M_{\gamma_i}^p = I, \tag{4.80}$$

where p is the order of the generator. However, not all of the roots of equation (4.80) lead to representations of the group, and in any case, how do we know when to stop? We could solve this equation for representations of any dimension, so when do we know that we have found all the inequivalent irreps?

For finite groups we can use Theorems 3.2, 3.3 and 3.4 to check whether we have all the irreps over $\mathbb{C}$ Since we are looking for the absolutely irreducible real representations, we will then need to check which of these are realisable over $\mathbb{R}$ Using D_4 as an example once more, we first need to find its order and conjugacy classes. D_4 has eight elements, so the order of the group is 8 To find the conjugacy classes, recall that two elements, h_1 and h_2, of D_4 are conjugate if there exists $\gamma \in D_4$ such that

$$h_1 = \gamma h_2 \gamma^{-1}, \tag{4.81}$$

and that conjugate elements have the same order The identity element is in a class on its own because

$$\gamma e \gamma^{-1} = e, \quad \forall \gamma \in D_4, \tag{4.82}$$

or equivalently because it is the only group element of order one. This is true for any group.

The rotations ρ and ρ^3, through $\pm\pi/2$ respectively, are the only group elements of order four, the remaining elements, other than the identity, having order two. This precludes any other elements being in the same class as ρ and ρ^3, so it remains to show that they are in the same class as each other. Checking that

$$(m\rho)\rho^3(m\rho)^{-1} = m\rho\rho^3\rho^{-1}m = m\rho^{-1}m = \rho mm = \rho \tag{4.83}$$

does it, so $\{\rho, \rho^3\}$ is one conjugacy class.

To work out the remaining classes we will use

$$\rho^n m = m\rho^{-n}, \quad \text{for } n = 1, 2, 3. \tag{4.84}$$

Thus we find

$$\rho^q m\rho^s\rho^{-q} = m\rho^{-q}\rho^{s-q} = m\rho^{s-2q} \tag{4.85}$$

for $q, s = 0, 1, 2, 3$, and

$$\begin{aligned}(m\rho^q)m\rho^s(m\rho^q)^{-1} &= m\rho^q m\rho^s\rho^{-q}m = mm\rho^{-q}\rho^s m\rho^q \\ &= m\rho^{q-s}\rho^q = m\rho^{2q-s}\end{aligned} \tag{4.86}$$

for $q, s = 0, 1, 2, 3$. Recall that $\rho^4 = e$. Choosing $s = 0$ shows that m and $m\rho^2$ form a class, and choosing $s = 1$ shows that $m\rho$ and $m\rho^3$ form another class. While it is easy to show, for example, that m and $m\rho^2$ are in the same class since

$$\rho m\rho^{-1} = m\rho^{-1}\rho^{-1} = m\rho^{-2} = m\rho^2, \tag{4.87}$$

we need to check all the other possibilities given in equations (4.85) and (4.86) to make sure that there are no other members of the class.

Similarly we have

$$\rho^q\rho^s\rho^{-q} = \rho^s \tag{4.88}$$

for $q, s = 0, 1, 2, 3$, and

$$\begin{aligned}(m\rho^q)\rho^s(m\rho^q)^{-1} &= m\rho^q\rho^s\rho^{-q}m = m\rho^s m \\ &= mm\rho^{-s} = \rho^{-s}\end{aligned} \tag{4.89}$$

for $q, s = 0, 1, 2, 3$, and so ρ^2 is in a class on its own, and as we deduced earlier, ρ and ρ^3 form another. Hence the conjugacy classes of $\mathcal{D}_4$ are $\{e\}$, $\{m, m\rho^2\}$, $\{m\rho, m\rho^3\}$, $\{\rho, \rho^3\}$, $\{\rho^2\}$, namely the identity, the reflections in axes joining midpoints of opposite sides, the diagonal reflections, the rotations through $\pm\pi/2$ and the rotation through π respectively. Note that each class contains elements whose actions are similar in a natural sense.

There are five conjugacy classes, so by Theorem 3.2 the number, n, of inequivalent irreducible representations over $\mathbb{C}$, is five.

Table 4.2

	e	$\{m, m\rho^2\}$	$\{m\rho, m\rho^3\}$	$\{\rho, \rho^3\}$	ρ^2
R_1	1	1	1	1	1
R_2	1	1	−1	−1	1
R_3	1	−1	−1	1	1
R_4	1	−1	1	−1	1
R_5	2				

Table 4.3

	e	$\{m, m\rho^2\}$	$\{m\rho, m\rho^3\}$	$\{\rho, \rho^3\}$	ρ^2
R_1	1	1	1	1	1
R_2	1	1	−1	−1	1
R_3	1	−1	−1	1	1
R_4	1	−1	1	−1	1
R_5	2	0	0	0	−2

By Theorem 3.3, we also know that

$$\sum_{i=1}^{5} d_i^2 = 8, \tag{4.90}$$

which has the unique solution $d_1 = d_2 = d_3 = d_4 = 1$ and $d_5 = 2$, since no d_i can be zero.

We know that the one-dimensional irreps can be constructed by setting $M_m = \pm 1$ and $M_\rho = \pm 1$ as discussed in Section 4.3.1 above. In the two-dimensional irrep, the identity element will be represented by the identity matrix, $M_e = I$, and so its character will be 2, the dimension of the irrep. Already we have the entries in the Table 4.2.

Now we can use the orthogonality theorems (Theorem 3.4) to complete the table. The first and last columns of the table must be orthogonal by equation (3.113), so ρ^2 must have character −2. Now the sum of the squares of the characters in the two-dimensional irrep, weighted by the number of elements in each class, must be 8, so the characters of all the remaining classes must be zero, and the completed character table is as given in Table 4.3.

Now we need to deduce the matrices of the irreps from their characters. The matrices of the one-dimensional irreps are identical to the characters. In this case

constructing the higher-dimensional irrep is easy, since we know that we have only one two-dimensional irrep, hence that it must be the natural representation and consequently straightforward to write down. Of course this only works because the natural representation turns out to be absolutely irreducible: it is easy to check that only multiples of the identity commute with all the matrices given in (4.54). Any irrep equivalent to the natural representation would also give the correct characters in the two-dimensional case. In this example, all the irreducible representations over $\mathbb{C}$ turn out to be real, so we know we have found all the absolutely irreducible real representations.

The whole procedure of finding the absolutely irreducible representations and the axial and maximal isotropy subgroups can be automated using a computer algebra package (see Matthews, 2004, for further details).

4.3.3 Weak symmetry breaking

Suppose that we find our experimental box is not perfectly square after all, but very slightly rectangular. We could go back and start the analysis all over again using the symmetry group, $\mathcal{D}_2$, of a rectangle. This turns out to have only one-dimensional irreps. What if, though, we think that in a truly square box the experiment is governed by the natural representation of $\mathcal{D}_4$ and we want to find out what effect the slight deviation from squareness has on the solutions? An easy way to do this is to break the square symmetry slightly in the two-dimensional $\mathcal{D}_4$ normal form equations (4.72). We don't have symmetry under rotation through $\pi/2$ any more, so the equations take the form

$$\begin{pmatrix} \mathrm{d}x_1/\mathrm{d}t \\ \mathrm{d}x_2/\mathrm{d}t \end{pmatrix} = \begin{pmatrix} \mu_1 x_1 \\ \mu_2 x_2 \end{pmatrix} - \begin{pmatrix} a_{11}x_1^3 \\ a_{12}x_2^3 \end{pmatrix} - \begin{pmatrix} a_{21}x_2^2 x_1 \\ a_{22}x_1^2 x_2 \end{pmatrix}, \tag{4.91}$$

satisfying only the reflections m and $m\rho^2$. This corresponds to using a two-dimensional reducible representation of $\mathcal{D}_2$. The symmetry breaking is weak, so we must have $\mu_1 \approx \mu_2 \approx \mu$, $a_{11} \approx a_{12} \approx a_1$ and $a_{21} \approx a_{22} \approx a_2$. Now to simplify things we will assume that $a_1 > 0$, and we can then rescale x_1 and x_2 to set $a_{11} = a_{12} = 1$. It turns out that unless $a_2 \approx 1$, there is no new behaviour gained by having a_{21} and a_{22} different, so we will set them both to a, and we can analyse the bifurcation using the equations

$$\begin{pmatrix} \mathrm{d}x_1/\mathrm{d}t \\ \mathrm{d}x_2/\mathrm{d}t \end{pmatrix} = \begin{pmatrix} \mu_1 x_1 \\ \mu_2 x_2 \end{pmatrix} - \begin{pmatrix} x_1^3 \\ x_2^3 \end{pmatrix} - a \begin{pmatrix} x_2^2 x_1 \\ x_1^2 x_2 \end{pmatrix} \tag{4.92}$$

The stationary solutions and their stabilities can be determined as functions of μ_1, μ_2 and a. In a real experiment, we would vary a control (bifurcation) parameter

related to both μ_1 and μ_2, so we would follow a curve in μ_1–μ_2 space as the bifurcation parameter increased.

Other types of symmetry can also be broken. Suppose the system had a preferred direction of rotation. We would retain symmetry under rotation through $\pi/2$, but there would no longer be any reflection symmetries. New cubic terms are allowed in the normal form which now corresponds to a two-dimensional representation of $\mathbb{Z}_4$ and looks like this:

$$\begin{pmatrix} \mathrm{d}x_1/\mathrm{d}t \\ \mathrm{d}x_2/\mathrm{d}t \end{pmatrix} = \mu \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - a_1 \begin{pmatrix} x_1^3 \\ x_2^3 \end{pmatrix} - a_2 \begin{pmatrix} x_2^2 x_1 \\ x_1^2 x_2 \end{pmatrix} - a_3 \begin{pmatrix} x_1^2 x_2 \\ -x_2^2 x_1 \end{pmatrix} - a_4 \begin{pmatrix} x_2^3 \\ -x_1^3 \end{pmatrix} \tag{4.93}$$

If the symmetry breaking were weak, we would expect a_3 and a_4 to be small. Again these equations can be analysed to find fixed points and their stabilities.

The advantage of investigating weak symmetry breaking, if it is appropriate for the situation you are looking at, is that you are just considering small deviations from a simple symmetric system that you have already analysed and understand well. Of course, if the symmetry breaking is really a large effect in the experiment you are trying to describe, then you must start again from the beginning using the appropriate reduced symmetry group.

4.3.4 Bifurcation on the surface of a sphere

The equivariant branching lemma is also useful for symmetry groups that contain continuous symmetries, such as the orthogonal group, $O(3)$, the symmetry group of a sphere. This group is the direct product of the special orthogonal group, $SO(3)$, of rotations of the sphere, and $\mathbb{Z}_2^c$, the group generated by the inversion, $\boldsymbol{x} \to -\boldsymbol{x}$, for $\boldsymbol{x} \in \mathbb{R}^n$, sometimes called the reflection through the origin, so we have

$$O(3) = SO(3) \times \mathbb{Z}_2^c. \tag{4.94}$$

The irreducible representations of $SO(3)$ can be described in terms of surface (spherical) harmonics, and from them the irreducible representations of $O(3)$ can be constructed by considering the two cases where $\mathbb{Z}_2^c$ acts as plus or minus the identity. This is all quite tractable (see Golubitsky, Stewart & Schaeffer, 1988, for the details), but a little long-winded, so we will investigate only one representation here.

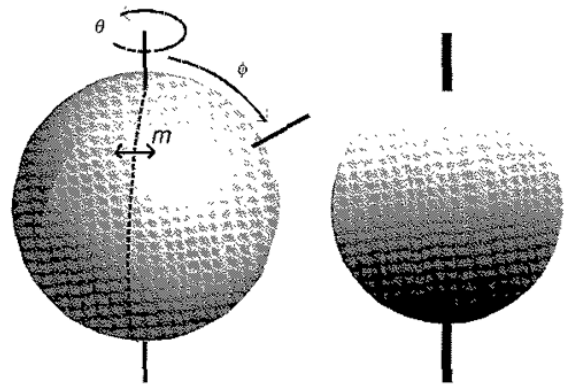


Fig 4 12 (a) The generators of the symmetry group of a sphere, and (b) a solution with isotropy subgroup $O(2)^-$, where the size of the solution amplitude is represented by the intensity of the grey shading

Coordinate geometry suggests the following three-dimensional representation, constructed from the matrices

$$M_\theta = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{4.95}$$

$$M_\phi = \begin{pmatrix} \cos\phi & 0 & -\sin\phi \\ 0 & 1 & 0 \\ \sin\phi & 0 & \cos\phi \end{pmatrix}, \tag{4.96}$$

$$M_m = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{4.97}$$

which correspond to a rotation of $0 \leq \theta < 2\pi$ about the z axis, a rotation of $0 \leq \phi < \pi$ about the y axis and a reflection in the yz plane respectively, where $x = (x, y, z)$ These symmetries are shown in Figure 4.12a. Now, clearly $O(3)$ is a compact Lie group and $\mathrm{Fix}(O(3)) = \{\mathbf{0}\}$ in this representation, which is in fact absolutely irreducible as you can verify by trying to find matrices that commute with all the generators and deciding that the only ones that do are multiples of the identity Imagine that we have a steady bifurcation problem with spherical symmetry that is governed by this representation of $O(3)$ We will assume that the form of the bifurcation problem satisfies the hypotheses of the equivariant branching lemma (Theorem 4 5), and so to determine the guaranteed primary solution branches at the bifurcation point we need to identify the isotropy subgroups with

one-dimensional fixed-point subspace We are working in the natural coordinate-geometry representation, so this amounts to identifying the largest subgroups that fix a line in the sphere. It is clear that each of these consists of rotations about a fixed axis and reflections in any plane that passes through this axis This is a group with $O(2)^-$ symmetry The minus sign of the $O(2)^-$ subgroup indicates that the order two generator is a true reflection, such as m, and does not lie in $SO(3)$ The existence of a primary branch of solutions with such symmetry is guaranteed by the generalised equivariant branching lemma A solution with this symmetry is illustrated in Figure 4 12b. The group $O(2)$ can also be generated by a circle of rotations about one axis and a rotation through π about a perpendicular axis, but this does not correspond to a guaranteed solution branch Note that the solution shown in Figure 4 12b has no such perpendicular π-rotation symmetry since the shading representing the solution amplitude varies from most to least intense along the axis of rotation.

The coordinate-geometry representation is associated with spherical harmonics of degree one, because the expressions for the Cartesian coordinates $x = r\cos\theta\sin\phi$, $y = r\sin\theta\sin\phi$, $z = r\cos\phi$ are exactly those spherical harmonics. At a general bifurcation from spherical symmetry, the eigenfunctions from which the bifurcating solutions are constructed are spherical harmonics of some degree Implicit in the analysis we have just been through is the assumption that the bifurcating solutions can be constructed from first order spherical harmonics In reality, we may see structures with much more complicated symmetry, and we would need to use higher-order spherical harmonics to describe them. There is no problem with this: we simply have to use a higher-dimensional representation to analyse the bifurcation problem. There are two irreducible representations of $O(3)$ associated with the spherical harmonics of degree l for every l. The equivariant branching lemma can be applied in each case: the predicted solution branches include those with icosahedral or octahedral symmetry, alongside axisymmetric solutions such as the one we have already found, and a whole host of other symmetry types besides (Matthews, 2003) Some of the patterns predicted are shown in Figure 4 13.

4.4 Hopf bifurcations with symmetry

Up to this point we have been discussing the effect of symmetry on steady bifurcations. Symmetric systems can also undergo Hopf bifurcations when complex conjugate eigenvalues cross the imaginary axis. In the standard Hopf bifurcation only one pair of complex conjugate eigenvalues crosses the imaginary axis at the bifurcation point, but in a symmetric system fixed points on the same orbit have the same stability properties, and so at a Hopf bifurcation with symmetry there are expected to be multiple pairs of complex conjugate eigenvalues crossing

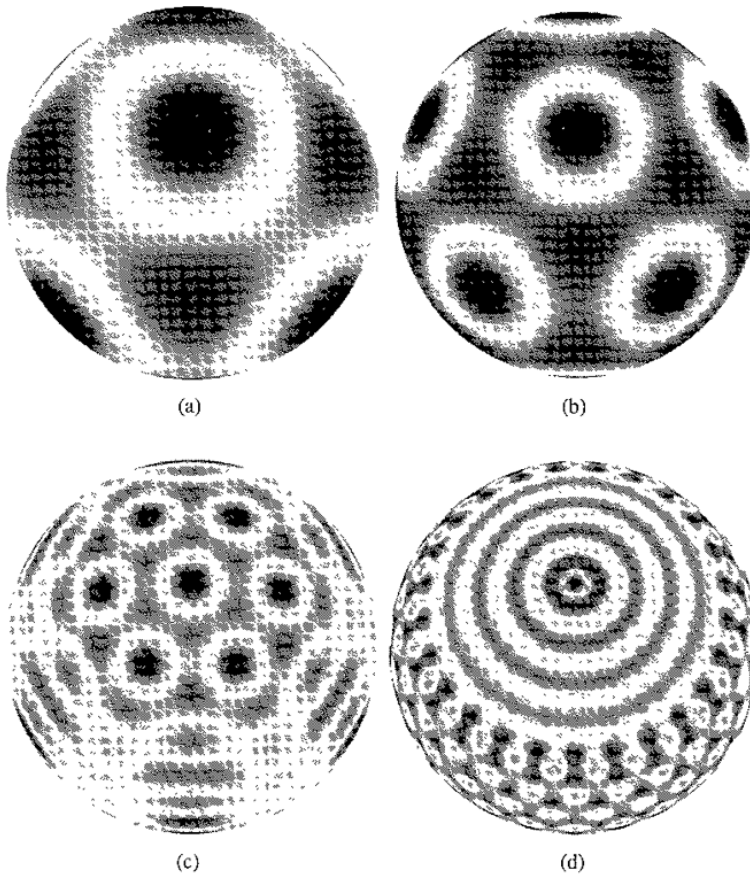


Fig 4.13 Some of the patterns predicted at a bifurcation with spherical symmetry for representations associated with spherical harmonics of degree l: (a) a state with cubic symmetry (O_h) for $l = 4$, (b) an icosahedral solution (symmetry group I_h) for $l = 6$, (c) a solution with hexagonal ($\mathcal{D}_6 \times \mathbb{Z}_2^c$) symmetry for $l = 14$ and (d) a solution with $\mathcal{D}_{25} \times \mathbb{Z}_2^c$ symmetry for $l = 30$. Pictures courtesy of and ©Paul Matthews, University of Nottingham, 2003

the imaginary axis at once. In this section we will discuss some of the main results and show how they can be applied in practice: if you would like to see the proofs they can be found in Golubitsky, Stewart and Schaeffer (1988).

As usual let us start with the equation

$$\frac{d\boldsymbol{x}}{dt} = f(\boldsymbol{x}, \mu), \tag{4.98}$$

where $\boldsymbol{x} \in \mathbb{R}^n$, $\mu \in \mathbb{R}$ and f is smooth, and which is equivariant under the action of a symmetry group, Γ, so that we have

$$\gamma f(\boldsymbol{x}, \mu) = f(\gamma \boldsymbol{x}, \mu), \quad \forall \gamma \in \Gamma. \tag{4.99}$$

We want a Γ-invariant equilibrium solution, $x = \mathbf{0}$, so we must have

$$f(\mathbf{0}, \mu) = \mathbf{0}, \quad \forall \mu \tag{4.100}$$

So far this is all the same as for the symmetric steady bifurcations we have just been looking at, but the difference is that for a Hopf bifurcation we want the Jacobian $Df|_{(\mathbf{0},0)}$ to have purely imaginary eigenvalues. Just as the generic action of Γ on the centre eigenspace is absolutely irreducible for a steady-state Γ-equivariant bifurcation, which forces the eigenvalues of the Jacobian to be real, it turns out that the imaginary eigenspace must be Γ-simple if $Df|_{(\mathbf{0},0)}$ is to have purely imaginary eigenvalues and a Hopf bifurcation is to be possible

A representation, W, of Γ is defined to be **Γ-simple** either if it is composed of two copies of an absolutely irreducible representation V so that $W \cong V \oplus V$, or if W is irreducible, but not absolutely irreducible. In either case, in suitable coordinates at the bifurcation point the Jacobian generically takes the form

$$Df|_{(\mathbf{0},0)} = J \equiv \begin{pmatrix} 0 & I_p \\ -I_p & 0 \end{pmatrix}, \tag{4.101}$$

where $p = n/2$, and the eigenvalues of $Df|_{(\mathbf{0},\mu)}$ are

$$\lambda_{\pm} = \sigma(\mu) \pm i\omega(\mu), \tag{4.102}$$

each of multiplicity p, where σ and ω are smooth functions of μ satisfying $\sigma(0) = 0$ and $\omega(0) = 1$. Note that since this implies that the eigenvalues at the bifurcation point are $\pm i$, you might have to rescale time to put the Jacobian in this form.

There is an equivalent of the equivariant branching lemma that we can use to predict the generic bifurcating solutions for Hopf bifurcations with symmetry, but before we can use it we need to introduce the idea of a spatiotemporal symmetry.

At Hopf bifurcations, we expect to see branches of periodic solutions, rather than the stationary solution branches that occur at steady-state bifurcations. Periodic solutions can have purely spatial symmetries just as stationary solutions do: if for some $\gamma \in \Gamma$ a periodic solution, $x(t)$, satisfies

$$\gamma x(t) = x(t), \quad \forall t, \tag{4.103}$$

then γ is a spatial symmetry of the solution. In addition, though, periodic solutions can have spatiotemporal symmetries: if $x(t)$ has period 2π in t and satisfies

$$(\gamma, \theta) \quad x(t) \equiv \gamma x(t + \theta) = x(t), \quad \forall t, \tag{4.104}$$

where $(\gamma, \theta) \in \Gamma \times S^1$ and S^1 is the circle group of phase shifts acting on the space of 2π-periodic functions, then (γ, θ) is a **spatiotemporal symmetry**. Equation

(4.104) says that performing a spatial action γ on $x(t)$ followed by a shift in time (phase shift) of θ leaves the solution unchanged. The symmetry consists of both spatial and temporal elements, hence its name. Of course, if θ is zero then the symmetry is purely spatial, so the set of all spatiotemporal symmetries includes the purely spatial ones, and we can write the isotropy subgroup of $x(t)$ as

$$\Sigma_{x(t)} = \{(\gamma, \theta) \in \Gamma \times S^1 : \gamma x(t+\theta) = x(t)\} \subset \Gamma \times S^1. \tag{4.105}$$

The action of the phase-shift symmetry can be deduced from the linearisation of equation (4.98) on the imaginary eigenspace, namely

$$\frac{dx}{dt} = Jx. \tag{4.106}$$

The general solution to this equation is $x(t) = e^{tJ}x(0)$. Now the action of the phase shift is

$$(e, \theta) \quad x(t) = x(t+\theta) = e^{(t+\theta)J}x(0) = e^{(\theta+t)J}x(0) = e^{\theta J}e^{tJ}x(0) = e^{\theta J}x(t) \tag{4.107}$$

In the case where the imaginary eigenspace is of the form $V \oplus V$, with V absolutely irreducible, we can use the definition of J given in equation (4.101) to find

$$\begin{aligned} e^{\theta J} &= \begin{pmatrix} I_p & 0 \\ 0 & I_p \end{pmatrix} + \begin{pmatrix} 0 & \theta I_p \\ -\theta I_p & 0 \end{pmatrix} \\ &\quad + \frac{1}{2!}\begin{pmatrix} -\theta^2 I_p & 0 \\ 0 & -\theta^2 I_p \end{pmatrix} + \frac{1}{3!}\begin{pmatrix} 0 & -\theta^3 I_p \\ \theta^3 I_p & 0 \end{pmatrix} + \cdots \qquad (4.108) \\ &= \begin{pmatrix} \cos\theta I_p & \sin\theta I_p \\ -\sin\theta I_p & \cos\theta I_p \end{pmatrix}. \qquad (4.109) \end{aligned}$$

We can deduce that the action of the phase shift on $V \oplus V$ is given by

$$(e, \theta) \quad (x_1, x_2) = [x_1|x_2]R_\theta, \tag{4.110}$$

where $(x_1, x_2) \in V \oplus V$, $[x_1|x_2]$ is the $p \times 2$ matrix formed by writing the x_1 and x_2 column vectors next to each other, and

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{4.111}$$

is the matrix for rotation through an angle θ.

It turns out that generically $\Gamma \times S^1$ acts nonabsolutely irreducibly on the Γ-simple imaginary eigenspace, and so there is an analogy between the role of the symmetry group $\Gamma \times S^1$ in the Hopf case and that of the symmetry group Γ in the

steady-state case. The equivariant Hopf theorem is the analogue of the equivariant branching lemma

Theorem 4.6 (Equivariant Hopf theorem) *Let*

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x, \mu), \quad x \in \mathbb{R}^n, \ \mu \in \mathbb{R} \tag{4.112}$$

be a Γ-equivariant Hopf bifurcation problem satisfying equations (4.99), (4.101) and (4.102) and also

$$\left.\frac{\mathrm{d}\sigma}{\mathrm{d}\mu}\right|_{\mu=0} \neq 0. \tag{4.113}$$

If Γ acts Γ-simply on $\mathbb{R}^n$ and $\Sigma \subset \Gamma \times S^1$ is an isotropy subgroup satisfying

$$\dim \mathrm{Fix}(\Sigma) = 2, \tag{4.114}$$

then there exists a unique branch of periodic solutions, with period close to 2π, bifurcating from the origin such that the isotropy subgroup of each solution is Σ

We can use the theorem to predict the symmetry of bifurcating solution branches, just as we did the equivariant branching lemma, the main difference being that now we need to identify the isotropy subgroups that have a two-dimensional fixed-point subspace. To see how this works, we will look at a Hopf bifurcation with $O(2)$ symmetry, which is common in pattern-forming systems.

4.4.1 Hopf bifurcation with $O(2)$ symmetry

$O(2)$ is the group of reflections and rotations of the plane: its standard action on $\mathbb{R}^2 \cong \mathbb{C}$ is generated by a rotation, $0 \leq \rho < 2\pi$, and a reflection, m, given by

$$\rho w = e^{i\rho} w, \tag{4.115}$$

$$mw = \bar{w}, \tag{4.116}$$

where $\bar{w}$ is the complex conjugate of w. This action is irreducible, so to get a Hopf bifurcation we will need to have two copies of it. If we take a point $(w_1, w_2) \in \mathbb{C} \oplus \mathbb{C}$, and apply the standard action of $O(2)$ we find

$$\rho(w_1, w_2) = (e^{i\rho} w_1, e^{i\rho} w_2), \tag{4.117}$$

$$m(w_1, w_2) = (\bar{w}_1, \bar{w}_2). \tag{4.118}$$

We know from equation (4.110) that the phase shift acts according to

$$\theta(w_1, w_2) = [w_1 | w_2] R_\theta, \tag{4.119}$$

where

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{4.120}$$

and so

$$\theta(w_1, w_2) = (w_1 \cos\theta + w_2 \sin\theta, -w_1 \sin\theta + w_2 \cos\theta) \tag{4.121}$$

Changing coordinates to

$$z_1 = \frac{1}{2}(\bar{w}_1 - i\bar{w}_2), \tag{4.122}$$

$$z_2 = \frac{1}{2}(w_1 - iw_2) \tag{4.123}$$

makes the action look neater:

$$\rho(z_1, z_2) = (e^{-i\rho}z_1, e^{i\rho}z_2), \tag{4.124}$$

$$m(z_1, z_2) = (z_2, z_1), \tag{4.125}$$

$$\theta(z_1, z_2) = (e^{i\theta}z_1, e^{i\theta}z_2) \tag{4.126}$$

The spatiotemporal symmetries belong to the group $O(2) \times S^1$, and group elements take the form (γ, θ), where $\gamma \in O(2)$ is generated by ρ and m. If γ is a pure rotation, this notation is potentially confusing, so just to emphasize, the element $(\pi/2, \pi/3)$, for example, represents a rotation through $\pi/2$ followed by a time shift of $\pi/3$.

Now we need to identify the isotropy subgroups and find out which ones have a two-dimensional fixed-point subspace. The symmetries $(0, 0)$ and (π, π) leave all points (z_1, z_2) unchanged, so they are trivial symmetries of all solutions: the first is the identity element, the second is a rotation through π together with a half-period phase shift. They generate the isotropy subgroup $\mathbb{Z}_2^c$ that fixes the whole four-dimensional space $\mathbb{C}^2 \cong \mathbb{R}^4$ spanned by (z_1, z_2). The whole group $O(2) \times S^1$ leaves only the origin $(0, 0)$ fixed, and so has zero-dimensional fixed-point subspace. The two remaining orbit types are the interesting ones:

(i) Solutions (z, z) are invariant under the reflection m. The solution maintains this purely spatial symmetry at all times, while oscillating periodically. The axis of reflection is $z \in \mathbb{R}$ and does not vary, so the solution can't rotate, only oscillate in place: since it can't travel, this type of solution is often referred to as a **standing wave**. The isotropy subgroup is $\mathbb{Z}_2 \times \mathbb{Z}_2^c = \{(0, 0), (\pi, \pi), (m, 0), (m, 0)(\pi, \pi)\}$.

There is a whole family of standing-wave solutions given by $(z, ze^{-i\rho})$ with reflection symmetry $m\rho$ for $0 \le \rho < 2\pi$. The axis of reflection, $z = Re^{-i\rho/2}$, $R \in \mathbb{R}$, is rotated relative to the real axis, and each solution has an isotropy subgroup conjugate to the one given above for the special case $\rho = 0$.

Table 4.4. *The solution branches at a Hopf bifurcation with $O(2)$ symmetry*

Branch	Orbit representative	Isotropy subgroup Σ	Fixed-point subspace	dim Fix(Σ)
Trivial solution	(0,0)	$O(2) \times S^1$	$\{0\}$	0
Rotating waves	$(a, 0), a > 0$	$\widetilde{SO(2)}$	$\{(z_1, 0)\}$	2
Standing waves	$(a, a), a > 0$	$\mathbb{Z}_2 \oplus \mathbb{Z}_2^c$	$\{(z_1, z_1)\}$	2
General points	$(a, b), a > b > 0$	$\mathbb{Z}_2^c$	$\{(z_1, z_2)\}$	4

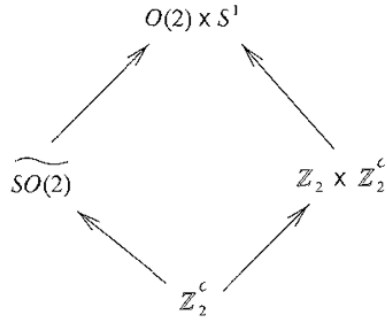


Fig. 4.14. Lattice of isotropy subgroups for the group $\Gamma = O(2) \times S^1$. Inclusion is shown by an arrow

(ii) Solutions $(z, 0)$ have isotropy subgroup $\widetilde{SO(2)} = \{(\theta, \theta) : 0 \leq \theta < 2\pi\}$ (where this serves as a definition of $\widetilde{SO(2)}$) For these solutions, a phase shift of θ in time is the same as a rotation through $-\theta$, so they rotate with time, and are known as **rotating waves** Waves that rotate in the opposite direction have the conjugate isotropy subgroup $\widetilde{SO(2)} = \{(-\theta, \theta)\}$ with fixed-point subspace $(0, z)$

In both these cases the fixed-point subspace is two-dimensional as $z \in \mathbb{C}$ so the solutions are specified by two independent coordinates, $Re(z)$ and $Im(z)$ The equivariant Hopf theorem says that generically a branch of periodic solutions with each of these isotropy subgroups will bifurcate from the origin at the Hopf bifurcation. The solution branches at a Hopf bifurcation with $O(2)$ symmetry are summarised in Table 4.4, and Figure 4.14 shows the isotropy lattice.

4.5 Heteroclinic cycles

Another way in which cyclic behaviour can occur after a bifurcation is through the formation of heteroclinic cycles A **heteroclinic cycle** consists of a set of fixed points, each solution being connected to the next by a forward-running trajectory in

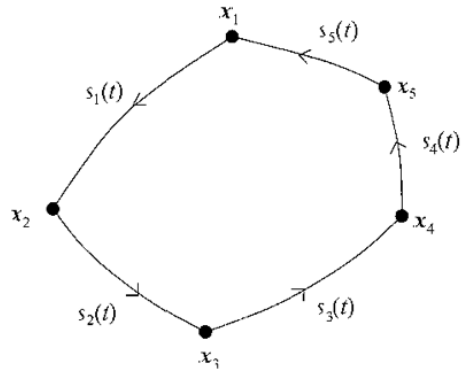


Fig. 4.15 A schematic diagram of a heteroclinic cycle with trajectories $s_1(t)$ to $s_5(t)$ running between fixed points x_1 to x_5

phase space to form a closed cycle. In other words if the equilibrium solutions are x_i for $i = 1, \ldots, n$, then there is a trajectory that leaves x_i and runs to x_{i+1}, where we identify x_{n+1} with x_1. The trajectory between one fixed point and the next is known as a **heteroclinic connection** and can be written $s_i(t)$, where $s_i(t) \to x_i$ as $t \to -\infty$ and $s_i(t) \to x_{i+1}$ as $t \to +\infty$. Figure 4.15 shows an example of a heteroclinic cycle.

Guckenheimer and Holmes (1988) showed that heteroclinic cycles in symmetric systems can be structurally stable. They took as their example the Busse–Heikes cycle (Busse & Heikes, 1980), which models the Küppers–Lortz instability (Küppers & Lortz, 1969) in rotating convection, described in more detail in Section 5.6. Busse and Heikes' equations are in fact identical to those derived by May and Leonard (1975) who described a cycle between three competing species in a population model.

Guckenheimer and Holmes' treatment of the Busse–Heikes cycle is very famous, so we shall follow it now to explain the notion of structural stability. Consider the group Γ acting on $\mathbb{R}^3$ and generated by a cyclic permutation of the coordinate axes given by

$$r = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \tag{4.127}$$

and a reflection in the yz plane given by

$$m = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.128}$$

In fact $\Gamma = \mathcal{T} \times \mathbb{Z}_2$, where $\mathcal{T}$ is the symmetry group of rotations of a regular tetrahedron. To cubic order the Γ-equivariant amplitude equations for $(x, y, z) \in \mathbb{R}^3$ are

$$\frac{dx}{dt} = x(\mu - ax^2 - by^2 - cz^2), \tag{4.129}$$

$$\frac{dy}{dt} = y(\mu - ay^2 - bz^2 - cx^2), \tag{4.130}$$

$$\frac{dz}{dt} = z(\mu - az^2 - bx^2 - cy^2), \tag{4.131}$$

where μ is a real bifurcation parameter, and a, b and c are real constants. It is clear from these equations that the coordinate planes $x = 0$, $y = 0$ and $z = 0$ are flow-invariant, as are the coordinate axes and the lines $x = \pm y = \pm z$.

The fixed points of this system are:

(i) $x = y = z = 0$;
(ii) $x^2 = \mu/a$, $y = z = 0$, and cyclic permutations of $\{x, y, z\}$;
(iii) $x = 0$, $y^2 = \mu(a - c)/(a^2 - bc)$, $z^2 = \mu(a - b)/(a^2 - bc)$, and cyclic permutations of $\{x, y, z\}$;
(iv) $x^2 = y^2 = z^2 = \mu/(a + b + c)$

We will assume that μ is positive, so that the fixed point at the origin has undergone a steady bifurcation and lost stability. We will also assume that $a > 0$ and $a + b + c > 0$, so that the fixed points on the coordinate axes and the lines $x = \pm y = \pm z$ exist, that $(a - c) > 0$ and $(a - b) < 0$, so that the fixed points off the coordinate axes in the coordinate planes do not exist, and that $b + c - 2a > 0$.

Taking the time derivative of $r(x, y, z) = x^2 + y^2 + z^2$ we find

$$\frac{1}{2}\frac{dr}{dt} = \left(x\frac{dx}{dt} + y\frac{dy}{dt} + z\frac{dz}{dt}\right), \tag{4.132}$$

$$= \mu(x^2 + y^2 + z^2) - a(x^4 + y^4 + z^4) \tag{4.133}$$

$$- (b + c)(x^2y^2 + y^2z^2 + z^2x^2), \tag{4.134}$$

$$= \mu(x^2 + y^2 + z^2) - a(x^2 + y^2 + z^2)^2 \tag{4.135}$$

$$- (b + c - 2a)(x^2y^2 + y^2z^2 + z^2x^2) \tag{4.136}$$

So we have $dr/dt < 0$ if $r \geq \mu/a$ and $r(x, y, z)$ decreases along trajectories while they lie outside or on the surface of the sphere $r = \mu/a$. Note the inequality

$$h(x, y, z) \equiv x^2y^2 + y^2z^2 + z^2x^2 - \frac{1}{3}(x^2 + y^2 + z^2)^2 \leq 0, \tag{4.137}$$

which you can check by finding the maximum of $h(x, y, z)$ and seeing that it is zero. The equality holds if and only if $x = y = z = 0$ or $x^2 = y^2 = z^2$.

Substituting into equation (4.136) gives

$$\frac{1}{2}\frac{dr}{dt} \geq \mu(x^2+y^2+z^2) - a(x^2+y^2+z^2)^2 \tag{4.138}$$

$$-\frac{1}{3}(b+c-2a)(x^2+y^2+z^2)^2, \tag{4.139}$$

$$= -\frac{1}{3}(a+b+c)r\left(r - \frac{3\mu}{(a+b+c)}\right), \tag{4.140}$$

with the equality only for $x = y = z = 0$ or $x^2 = y^2 = z^2$. This tells us that $dr/dt > 0$ for $0 < r < 3\mu/(a+b+c)$, so r increases along all nonzero trajectories inside the sphere, S, defined by $r \leq 3\mu/(a+b+c)$, and that apart from the fixed points $x = \pm y = \pm z = \pm\sqrt{\mu/(a+b+c)}$ all trajectories starting on the surface of S leave S transversely.

Thus all nonzero trajectories eventually enter the closed spherical shell, A, given by

$$\frac{3\mu}{(a+b+c)} \leq x^2+y^2+z^2 \leq \frac{\mu}{a}. \tag{4.141}$$

Now taking the time derivative of $g(x, y, z) = x^2y^2z^2$ we find that

$$\frac{dg}{dt} = 2xyz\left(\frac{dx}{dt}yz + x\frac{dy}{dt}z + xy\frac{dz}{dt}\right), \tag{4.142}$$

$$= 2x^2y^2z^2(3\mu - (a+b+c)r), \tag{4.143}$$

so we have $dg/dt \leq 0$ for $r \geq 3\mu/(a+b+c)$, with equality if and only if (x, y, z) lies on one of the coordinate planes, where $g(x, y, z) = 0$, or the surface of the sphere, S, where $r = 3\mu/(a+b+c)$. This means that $g(x, y, z)$ decreases along trajectories outside the sphere, S, and hence decreases along trajectories inside the spherical shell, A.

Taken together the behaviour of r and g on trajectories tell us that all trajectories other than those on the invariant lines $x = \pm y = \pm z$ are attracted towards the coordinate planes.

Let us consider the coordinate plane $z = 0$, where the amplitude equations take the form

$$\frac{dx}{dt} = x(\mu - ax^2 - by^2), \tag{4.144}$$

$$\frac{dy}{dt} = y(\mu - ay^2 - cx^2). \tag{4.145}$$

The fixed points in this plane for the range of parameter values we have chosen are $(0, 0, 0)$ which is an unstable node, $(\pm\sqrt{\mu/a}, 0, 0)$ and $(0, \pm\sqrt{\mu/a}, 0)$. The Poincaré–Bendixson theorem implies that any bounded trajectory in an invariant

plane must approach either a fixed point or a periodic orbit. The statement and proof of this theorem can be found in Glendinning (1994). Using the Poincaré index (details of which can once again be found in Glendinning's book) it can be shown that any periodic orbit must enclose a fixed point, but here the only fixed points lie on the coordinate axes, and since the coordinate axes are invariant no trajectory can cross them, so no periodic orbit can exist. The trajectories must end up at fixed points then, so we need to work out which ones are stable and which unstable.

Perturbing the fixed points $(\pm\sqrt{\mu/a}, 0, 0)$ in the invariant plane $z = 0$ by setting $x = \sqrt{\mu/a} + \xi$ and $y = \eta$, where $|\xi|, |\eta| \ll 1$, and linearising gives

$$\frac{d\xi}{dt} = -2\mu\xi, \tag{4.146}$$

$$\frac{d\eta}{dt} = \frac{\mu}{a}(a - c)\eta, \tag{4.147}$$

so they are saddle points since $\mu > 0$ and $(a - c) > 0$. Now perturbing the fixed points $(0, \pm\sqrt{\mu/a}, 0)$ by setting $x = \xi$ and $y = \sqrt{\mu/a} + \eta$, where $|\xi|, |\eta| \ll 1$, and linearising gives

$$\frac{d\xi}{dt} = \frac{\mu}{a}(a - b)\xi, \tag{4.148}$$

$$\frac{d\eta}{dt} = -2\mu\eta, \tag{4.149}$$

so they are stable nodes in the plane $z = 0$ since $\mu > 0$ and $(a - b) < 0$. This means that there must be a trajectory joining the saddle point $(\sqrt{\mu/a}, 0, 0)$ to the stable node $(0, \sqrt{\mu/a}, 0)$ in the quarter-plane $x > 0$, $y > 0$, $z = 0$. However since the equations are equivariant under cyclic permutations of the coordinate axes, if we now look at the plane $x = 0$, the point $(0, \sqrt{\mu/a}, 0)$ becomes a saddle point in the new invariant plane, while the point $(0, 0, \sqrt{\mu/a})$ is a stable node, and there is a trajectory joining them in the quarter plane $y > 0$, $z > 0$, $x = 0$. The cycle is completed by a trajectory in $z > 0$, $x > 0$, $y = 0$ joining $(0, 0, \sqrt{\mu/a})$ to $(\sqrt{\mu/a}, 0, 0)$. This heteroclinic cycle is shown in Figure 4.16. Each fixed point in the cycle is unstable in a direction towards the next point on the cycle, but stable in the transverse directions. Trajectories are attracted to the coordinate planes, so they will approach each fixed point in the stable transverse directions and then head off in the unstable direction of the heteroclinic orbit. The heteroclinic orbit thus attracts trajectories, and is said to be stable.

There are seven other heteroclinic cycles related to this one by reflection in the coordinate planes: they lie on the group orbit of the first cycle. All the trajectories that do not lie on the invariant lines $x = \pm y = \pm z$ or on the coordinate planes are

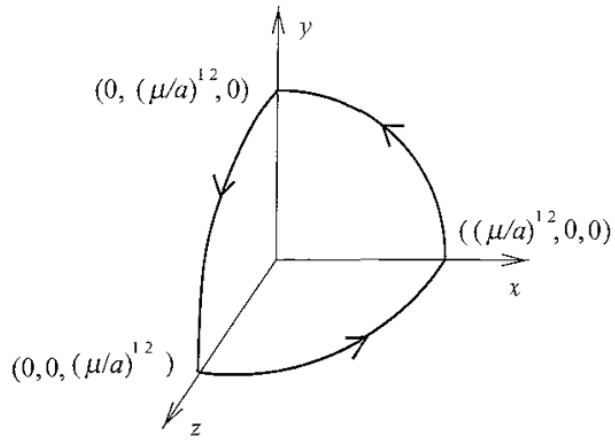


Fig 4.16. The Busse–Heikes heteroclinic cycle in the system of equations (4 129)–(4.131)

attracted towards the union of these heteroclinic cycles. A generic trajectory comes close to each of the fixed points in the cycle in turn, always moving closer to the coordinate planes, but never crossing them, so any one trajectory is attracted to only one cycle As you get closer to a fixed point, the movement along the trajectory gets slower because $|(dx/dt, dy/dt, dz/dt)|$ is close to zero, so as the trajectory approaches the heteroclinic cycle it spends longer near each fixed point in turn.

If the system is perturbed, then as long as the Γ-equivariance is not broken, it will still have to take the form given in equations (4.129)–(4 130), though the parameters μ, a, b and c will change a bit The properties that lead to the existence of the heteroclinic cycle hold in an open region of parameter space however, so the cycle will persist under small perturbations of the governing equations, though the equilibria on the coordinates axes will shift a bit, and the heteroclinic connections will be slightly deformed In particular the coordinate planes will remain invariant, and the saddle-sink connections in these planes will not break under small perturbations. We say that the cycle is **structurally stable** The idea of structural stability and its precise technical definition are discussed in more depth in Glendinning (1994) and Hirsch and Smale (1974).

The fixed points of the heteroclinic cycle have isotropy subgroups $\mathbb{Z}_2^2$ with one-dimensional fixed-point subspaces, the coordinate axes, while the heteroclinic connections lie in two-dimensional fixed-point subspaces corresponding to isotropy subgroups $\mathbb{Z}_2$, in other words the flow-invariant coordinate planes The relevant part of the isotropy lattice is shown in Figure 4 17 Melbourne, Chossat and Golubitsky (1989) pointed out that for a given action of a group, Γ, a characteristic

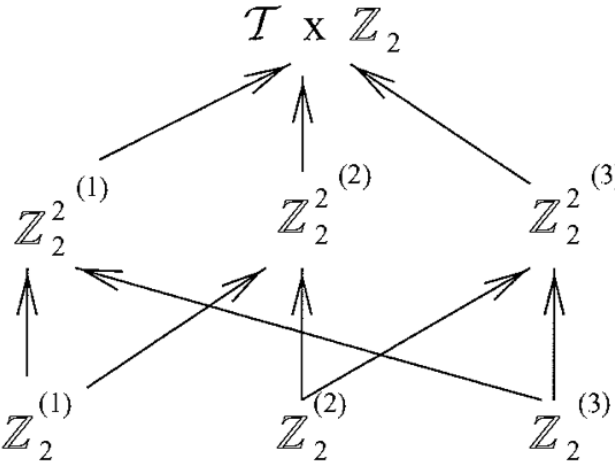


Fig. 4.17. The relevant part of the isotropy lattice for the Busse–Heikes cycle example. The isotropy subgroups of the three fixed points in the heteroclinic cycle are $\mathbb{Z}_2^{2(1)}$, $\mathbb{Z}_2^{2(2)}$ and $\mathbb{Z}_2^{2(3)}$, while the isotropy subgroups of the three heteroclinic connections are $\mathbb{Z}_2^{(1)}$, $\mathbb{Z}_2^{(2)}$ and $\mathbb{Z}_2^{(3)}$. Inclusion is shown by an arrow.

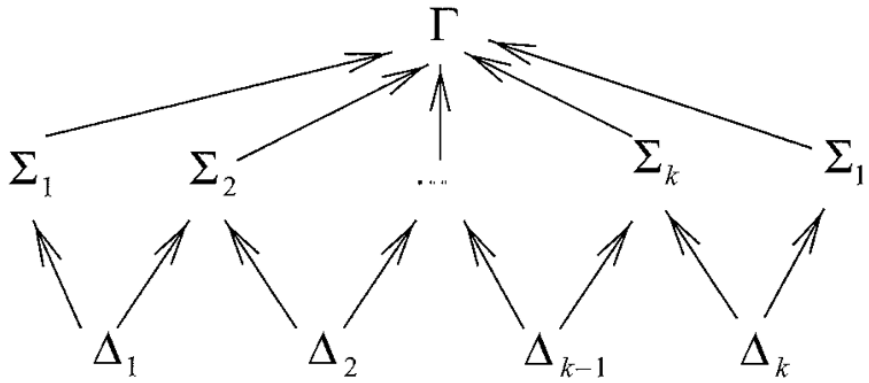


Fig. 4.18. Pattern in the isotropy lattice of a group, Γ, suggesting the existence of a structurally stable heteroclinic cycle. The isotropy subgroups, Σ_i, of the fixed points of the possible heteroclinic cycle have a one-dimensional fixed-point subspace, while the isotropy subgroups, Δ_j, of the possible heteroclinic connections have a two-dimensional fixed-point subspace. After Melbourne, Chossat and Golubitsky (1989).

pattern in the isotropy lattice, of the form shown in Figure 4.18, a generalisation of the pattern seen in the Busse–Heikes cycle example, suggests the existence of a structurally stable heteroclinic cycle. In both Figure 4.17 and Figure 4.18 conjugate isotropy subgroups appear independently in the isotropy lattice, contrary to convention, to clarify the possible heteroclinic connections.

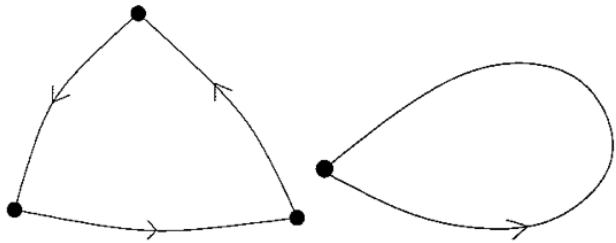


Fig. 4.19. (a) A homoclinic cycle consists of fixed points on the same group orbit, linked by trajectories, while (b) a homoclinic orbit consists of a trajectory linking a single fixed point back to itself.

A heteroclinic cycle consisting of fixed points that all lie on the same group orbit, as in the Busse–Heikes cycle, is often known as a **homoclinic cycle**. This is quite a common situation, but there are many other systems where the heteroclinic cycles connect fixed points of different types, lying on different group orbits. A homoclinic cycle is not to be confused with a **homoclinic orbit**, where a single trajectory loops round to connect a fixed point back to itself (see Figure 4.19).

For any given heteroclinic cycle in a Γ-equivariant system consisting of fixed points, $\boldsymbol{x}_i$, and connections, $s_i(t)$, for $i = 1, \ldots, n$, there is a group orbit of related heteroclinic cycles with fixed points $\gamma \boldsymbol{x}_i$ and connections $\gamma s_i(t)$, $\forall \gamma \in \Gamma$. It is conventional to identify cycles on the same group orbit, and hence to refer to the whole group orbit of cycles as one cycle. If a system contains more than one such cycle, in other words if there are cycles that are not related by symmetry, then if the unrelated cycles intersect they form a **heteroclinic network**. A heteroclinic network is a union of heteroclinic cycles that cannot be broken down into smaller disjoint networks or cycles. Two cycles intersect if they have in common at least two fixed points joined by a heteroclinic connection. At the point where the cycles diverge any trajectory close to the cycles has a 'choice' over which cycle to follow, so it is not clear that cyclic behaviour will result. For example, the trajectory could follow cycle A twice and then choose cycle B once, followed by cycle A eleven times and then cycle B eight times, or it could follow cycle A every time: the possible dynamics are very complicated and will depend on the details of the particular system under study. Kirk and Silber (1994) give an example of a heteroclinic network consisting of two cycles in a system with $\mathbb{Z}_2^4$ symmetry.

The definition of a heteroclinic cycle can be extended to include cycles between periodic orbits (Melbourne, Chossat & Golubitsky 1989) or chaotic sets, where the phenomenon is known as **cycling chaos** (Dellnitz *et al.* 1995).

Exercises

4.1 Consider the system

$$\frac{dx}{dt} = f(x, y),$$
$$\frac{dy}{dt} = g(x, y),$$

under the symmetries

$$\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} -x \\ y \end{pmatrix}, \quad \begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} -y \\ x \end{pmatrix}.$$

Show that these symmetries imply that f is odd in x and even in y, while g is odd in y and even in x.

4.2 Consider the system

$$\frac{dx}{dt} = f(x, y, z),$$
$$\frac{dy}{dt} = g(x, y, z),$$
$$\frac{dz}{dt} = h(x, y, z),$$

under the symmetries

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \to \begin{pmatrix} y \\ z \\ x \end{pmatrix}, \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \to \begin{pmatrix} x \\ z \\ y \end{pmatrix} \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \to \begin{pmatrix} -x \\ -y \\ -z \end{pmatrix}.$$

Show that these symmetries imply

$$h(y, z, x) = g(z, x, y) = f(x, y, z),$$
$$f(x, y, z) = f(x, z, y),$$
$$f(x, y, z) = -f(-x, -y, -z),$$

and hence write down the series expansion for $f(x, y, z)$, $g(x, y, z)$ and $h(x, y, z)$ for small $|x|$, $|y|$ and $|z|$ to quadratic order.

4.3 Find all the absolutely irreducible real representations of $\mathcal{D}_3$, the symmetry group of an equilateral triangle, and hence work out all the possible solutions that are guaranteed at a steady bifurcation with $\mathcal{D}_3$ symmetry under one or other of these representations. Draw examples of the eigenmodes in an appropriate triangular box, and work out the relevant normal form equations.

4.4 Repeat the previous exercise for $\mathcal{D}_6$, the symmetry group of a regular hexagon, this time drawing the eigenmodes in a hexagonal box.

4.5 Analyse the Hopf bifurcation with $\mathbb{Z}_2$ symmetry where the action of the group $\mathbb{Z}_2 \times S^1$ is given by

$$(e, \theta)z = e^{i\theta} z, \quad (m, 0)z = -z,$$

for $z \in \mathbb{C} \cong \mathbb{R}^2$, where (e, θ) is the phase shift and $(m, 0)$ is the reflection. What is the spatiotemporal symmetry of the guaranteed solutions?

4.6 Repeat the previous exercise for $\mathbb{Z}_4$ symmetry where the action of the group $\mathbb{Z}_4 \times S^1$ is given by

$$(e, \theta)z = e^{i\theta}z, \quad (\rho, 0)z = -z,$$

for $z \in \mathbb{C} \cong \mathbb{R}^2$, where (e, θ) is the phase shift and $(\rho, 0)$ is the rotation through $\pi/2$. What is the purely spatial symmetry of the guaranteed solutions? How do they evolve over time?

4.7 Repeat the previous exercise for $\mathbb{Z}_9$ symmetry where the action of the group $\mathbb{Z}_9 \times S^1$ is given by

$$(e, \theta)z = e^{i\theta}z, \quad (\rho\ 0)z = e^{2\pi i/3}z,$$

for $z \in \mathbb{C} \cong \mathbb{R}^2$, where (e, θ) is the phase shift and $(\rho, 0)$ is the rotation through $2\pi/9$. In what important respect does the representation of the group differ in this case from those used in the previous two exercises? What is the purely spatial symmetry of the guaranteed solutions in this case? How do they evolve over time?

Extension: deduce the behaviour for the general case of Hopf bifurcation with $\mathbb{Z}_n$ symmetry.

4.6 Appendix: Proofs

This section contains the proofs of Lemma 4.2 and Theorem 4.3 (the trace formula).

Lemma 4.2 *Let* $\mathrm{Fix}(\Gamma) = \{\mathbf{0}\}$, *and let* $\Sigma \subsetneq \Gamma$ *be a subgroup. Then* Σ *is a maximal isotropy subgroup of* Γ *if and only if*

$$\dim \mathrm{Fix}(\Sigma) > 0, \tag{A4.1}$$

$$\dim \mathrm{Fix}(\Delta) = 0 \quad \textit{for every closed subgroup } \Delta \supsetneq \Sigma \tag{A4.2}$$

Proof To prove the forwards implication, assume that Σ is a maximal isotropy subgroup. Then Σ must fix some vector x by the definition of an isotropy subgroup, and since $\mathrm{Fix}(\Gamma) = \{\mathbf{0}\}$ and $\Sigma \subsetneq \Gamma$ we must have x nonzero, and hence equation (A4.1) holds. Suppose Δ fixes a vector x, where $\Sigma \subsetneq \Delta$. Then we have $\Sigma \subsetneq \Delta \subset \Sigma_x$, where Σ_x is the isotropy subgroup of x. Now since Σ is maximal, we must have $\Sigma_x = \Gamma$, and since $\mathrm{Fix}(\Gamma) = \{\mathbf{0}\}$ we must also have $x = \mathbf{0}$. So the only vector fixed by Δ is zero, and we have proved equation (A4.2) $\dim \mathrm{Fix}(\Delta) = 0$.

To prove the reverse implication, assume that Σ satisfies equation (A4.1). Then Σ must fix some nonzero vector x, and the isotropy subgroup Σ_x of that point x contains Σ, so that $\Sigma_x \supset \Sigma$. Now Σ_x is closed because it is an isotropy subgroup, so if we had $\Sigma_x \neq \Sigma$ then by equation (A4.2) we would have $\dim \mathrm{Fix}(\Sigma_x) = 0$,

and hence $\Sigma_x = \Gamma$, but then we would also have $x = \mathbf{0}$, which is a contradiction. Hence we must have $\Sigma_x = \Sigma$ and so Σ is an isotropy subgroup. If Σ is not maximal, then there exists an isotropy subgroup Δ such that $\Sigma \subsetneq \Delta \subsetneq \Gamma$. Now Δ must fix some nonzero vector y, and so using the same argument as before we find that $\dim \mathrm{Fix}(\Delta) = 0$ and so $y = \mathbf{0}$, which is a contradiction. Therefore Σ must be maximal □

Theorem 4.3 (Trace formula) *Let Γ be a finite group, and let $\Sigma \subset \Gamma$ be a subgroup. Then the dimension of the fixed-point subspace of Σ is given by*

$$\dim(\mathrm{Fix}(\Sigma)) = \frac{1}{|\Sigma|} \sum_{\sigma \in \Sigma} tr(\sigma). \tag{A4.3}$$

Proof Define a linear map $A : V \to V$ by

$$Ax = \frac{1}{|\Sigma|} \sum_{\sigma \in \Sigma} \sigma x, \quad \forall x \in V \tag{A4.4}$$

Now square A to get

$$A^2 = \frac{1}{|\Sigma|^2} \sum_{\sigma \in \Sigma} \sigma \sum_{\sigma' \in \Sigma} \sigma' = \frac{1}{|\Sigma|^2} \sum_{\sigma \in \Sigma, \sigma' \in \Sigma} \sigma\sigma' \tag{A4.5}$$

$$= \frac{1}{|\Sigma|^2} |\Sigma| \sum_{\sigma \in \Sigma} \sigma = \frac{1}{|\Sigma|} \sum_{\sigma \in \Sigma} \sigma = A, \tag{A4.6}$$

using $\sigma\Sigma = \Sigma, \forall \sigma \in \Sigma$. We can use this result to see that

$$A|\mathrm{Im}\ A = Id, \tag{A4.7}$$

where Id is the identity mapping, and

$$\mathrm{Im}A = \{x \in V : x = Ay, \ y \in V\} \tag{A4.8}$$

is the image of A, by taking $x = Ay \in \mathrm{Im}\ A$ and applying A to find

$$Ax = A^2 y = Ay = x \tag{A4.9}$$

We know that $\dim(\ker A) + \dim(\mathrm{Im}A) = \dim V$ by the rank-nullity theorem for linear mappings (for details, see Cohn, 1982, who calls it the rank formula), and we also know that $\ker A \cap \mathrm{Im}\ A = \{\mathbf{0}\}$ since if $x \in \mathrm{Im}\ A$, then $Ax = x$ by equation (A4.7), and if $x \in \ker A$ then $Ax = \mathbf{0}$ and so we have $x = \mathbf{0}$. Thus we must have

$$V = \ker A \oplus \mathrm{Im}\ A \tag{A4.10}$$

Equations (A4.7) and (A4.10) together tell us that

$$tr(A) = \dim(\mathrm{Im}A). \tag{A4.11}$$

Now we will show that $\mathrm{Fix}(\Sigma) = \mathrm{Im}\ A$, which will complete the proof. First we have

$$Ax = \frac{1}{|\Sigma|}\sum_{\sigma\in\Sigma}\sigma x = \frac{1}{|\Sigma|}\sum_{\sigma\in\Sigma}x = \frac{1}{|\Sigma|}|\Sigma|x = x, \quad \forall x \in \mathrm{Fix}\ (\Sigma) \tag{A4.12}$$

which shows that $x \in \mathrm{Im}\ A,\ \forall x \in \mathrm{Fix}\ (\Sigma)$. Conversely if $x = Ay \in \mathrm{Im}\ A$, then

$$\sigma' x = \sigma' A y = \sigma' \frac{1}{|\Sigma|}\sum_{\sigma\in\Sigma}\sigma y = \frac{1}{|\Sigma|}\sum_{\sigma\in\Sigma}\sigma'\sigma y \tag{A4.13}$$

$$= \frac{1}{|\Sigma|}\sum_{\sigma\in\Sigma}\sigma y = Ay = x, \quad \forall\sigma' \in \Sigma \tag{A4.14}$$

since $\sigma'\sigma \in \Sigma$, for all σ' and σ in Σ, and so $x \in \mathrm{Fix}(\Sigma),\ \forall x \in \mathrm{Im}\ A$. Thus $\mathrm{Fix}(\Sigma) = \mathrm{Im}\ A$ and so the final result

$$\dim(\mathrm{Fix}(\Sigma)) = \frac{1}{|\Sigma|}\sum_{\sigma\in\Sigma} tr(\sigma). \tag{A4.15}$$

follows from equations (A4.4) and (A4.11). □

5

Simple lattice patterns

Lattice patterns crop up quite regularly in nature and experiments: they are structures that repeat more or less regularly in space many times over. In convection, the archetypal pattern-forming scenario, just about any lattice pattern you can think of can be produced by choosing the experimental set-up carefully. Typically you will see periodic stripes (known in convection as rolls) or hexagons, but squares, triangles and other more exotic varieties are also possible. A huge amount of research on pattern formation has been driven by the desire to understand convection patterns, so there has been a great emphasis on methods for analysing spatially periodic patterns. In the next two chapters we will apply the ideas of bifurcation with symmetry to lattice patterns, and in Chapter 7 we will follow up by exploring the methods we can use when the spatial periodicity is broken slightly.

5.1 Lattices and lattice patterns

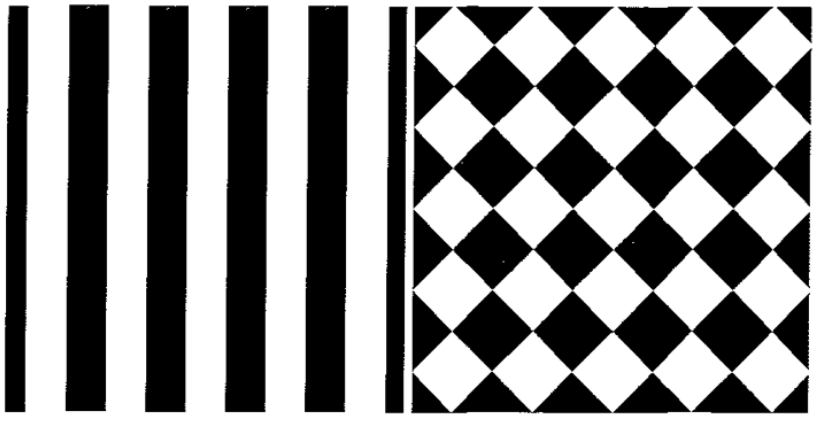
This whole chapter is about lattice patterns, so it is important to keep in mind what they look like. The common lattice patterns of stripes, squares and hexagons are shown in Figure 5.1.

A planar lattice, $\mathcal{L}$, is generated by two linearly independent vectors $\boldsymbol{l}_1$ and $\boldsymbol{l}_2 \in \mathbb{R}^2$ and is defined by

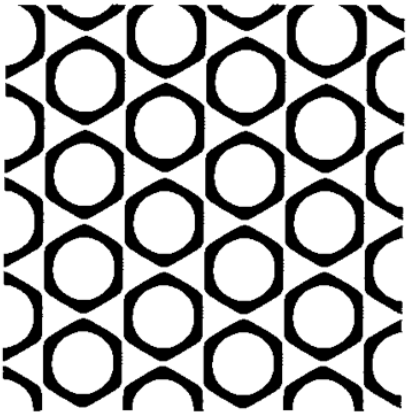
$$\mathcal{L} = \{n_1\boldsymbol{l}_1 + n_2\boldsymbol{l}_2 : n_1, n_2 \in \mathbb{Z}\} \tag{5.1}$$

The notation $\mathcal{L}$ is also used for the group of discrete translations that preserve the lattice. We will often want to characterise a lattice in terms of the dual lattice, $\mathcal{L}^*$, generated by two linearly independent wavevectors $\boldsymbol{k}_1$ and $\boldsymbol{k}_2 \in \mathbb{R}^2$ satisfying $\boldsymbol{k}_i \cdot \boldsymbol{l}_j = 2\pi\delta_{ij}$ for $i = 1, 2$, $j = 1, 2$. The dual lattice is defined by

$$\mathcal{L}^* = \{n_1\boldsymbol{k}_1 + n_2\boldsymbol{k}_2 : n_1, n_2 \in \mathbb{Z}\} \tag{5.2}$$



(a) (b)



(c)

Fig 5.1 Examples of some of the common lattice patterns: (a) stripes or rolls, (b) squares and (c) hexagons. These are filled contour plots of $u = \sum_j (e^{ik_j \cdot x} + e^{-ik_j \cdot x})$ for (a) $k_1 = (1, 0)$, (b) $k_1 = (1, 0), k_2 = (0, 1)$, and (c) $k_1 = (1, 0), k_2 = (-1/2, \sqrt{3}/2), k_3 = (-1/2, -\sqrt{3}/2)$, and so are purely linear superpositions of stripe patterns at various angles to each other For the stripes and squares, $u > 0$ regions are shown in white and $u < 0$ in black For the hexagons the central white region is $u > 0$, the black region is $-1\ 95 < u \leq 0$, and the outer white region is $u \leq -1\ 95$ In experiments there are usually some higher harmonics present, leading the patterns to look a little different even when the symmetries are the same

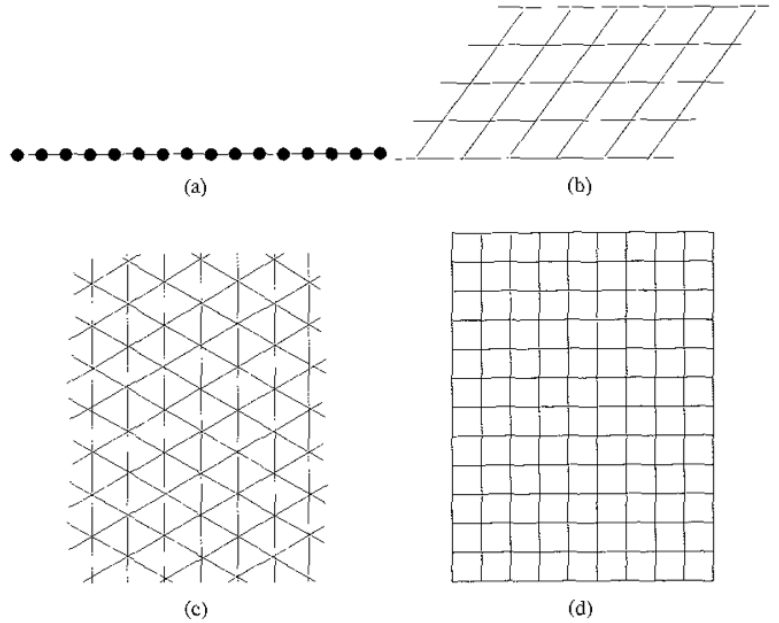


Fig 5 2 Dual lattices in one and two dimensions: (a) one-dimensional lattice, (b) rhombi ($|\boldsymbol{k}_1| = |\boldsymbol{k}_2| = k$, $\boldsymbol{k}_1 \cdot \boldsymbol{k}_2 \neq 0, k^2\cos(\pi/3), k^2\cos(2\pi/3)$), (c) hexagons ($|\boldsymbol{k}_1| = |\boldsymbol{k}_2| = k$, $\boldsymbol{k}_1 \cdot \boldsymbol{k}_2 = k^2\cos(2\pi/3)$) and (d) squares ($|\boldsymbol{k}_1| = |\boldsymbol{k}_2|$, $\boldsymbol{k}_1 \cdot \boldsymbol{k}_2 = 0$).

A planar lattice pattern is a function $u(\boldsymbol{x}, t)$ of a spatial variable $\boldsymbol{x} \in \mathbb{R}^2$ and time $t \in \mathbb{R}$ that is periodic on $\mathcal{L}$ so that

$$u(\boldsymbol{x} + \boldsymbol{l}, t) = u(\boldsymbol{x}, t) \quad \forall\, \boldsymbol{l} \in \mathcal{L} \tag{5.3}$$

In general an $\mathcal{L}$-periodic real function $u(\boldsymbol{x}, t)$ can be expressed as a sum of Fourier modes that lie on the dual lattice in the form

$$u(\boldsymbol{x}, t) = \sum_{k \in \mathcal{L}^*} z_k(t) e^{i\boldsymbol{k}\cdot\boldsymbol{x}} + c.c., \tag{5.4}$$

where $c.c.$ stands for the complex conjugate.

The two-dimensional periodic dual lattices that are most common in pattern formation have $|\boldsymbol{k}_1| = |\boldsymbol{k}_2|$. These are the square, rhombic and hexagonal lattices shown in Figure 5.2 alongside the one-dimensional periodic lattice

5.2 Bifurcations on a lattice

Pattern-forming systems are generally governed by a set of partial differential equations, so let us start by considering the equation

$$\frac{\partial u(\boldsymbol{x}, t)}{\partial t} = f(u(\boldsymbol{x}, t), \mu), \tag{5.5}$$

where f is a smooth nonlinear operator, $\mu \in \mathbb{R}$ is a bifurcation parameter, and u is a function of a spatial variable $\boldsymbol{x} \in \mathbb{R}^2$ and time $t \in \mathbb{R}$. For simplicity we will assume $u(\boldsymbol{x}, t) \in \mathbb{R}$, but it is possible to extend the method to $u(\boldsymbol{x}, t) \in \mathbb{R}^n$, so we can deal with a system that has several dependent variables and is described by more than one partial differential equation. The solution, $u(\boldsymbol{x}, t)$, represents a 'marker' quantity in the pattern-forming system, for example the vertical fluid velocity, the fluid density or temperature in convection, or the concentration of a product in a chemical reaction.

Patterns often occur in systems that have approximate Euclidean symmetry, namely symmetry under all rotations, reflections and translations of the plane. For example, if convection takes place in a uniform thin layer that is heated uniformly from below, then far enough away from the sidewalls of the container the experimental set-up has Euclidean symmetry, more or less, because it is the same in all directions (isotropic) and in all places (homogeneous). To have true Euclidean symmetry the experiment should take place on an infinite plane otherwise translations would move out of the experimental box at the boundaries. Of course the experimenters could decide to break the Euclidean symmetry by tilting the layer, so that the direction of tilt is different from all others and the isotropy is broken, or by heating the layer more intensely in one spot, so that the homogeneity is broken. The techniques we will use can be adapted to deal with these situations, but for now we will assume that equation (5.5) has Euclidean symmetry and that the solution evolves on the infinite plane. In this case f is equivariant with respect to E(2), the Euclidean group, and we have

$$\gamma f(u, \mu) = f(\gamma u, \mu), \quad \forall \gamma \in \mathrm{E}(2) \tag{5.6}$$

We will also assume that for all values of μ there is a time-independent solution to (5.5), invariant under the symmetries of the plane. This will be a spatially uniform stationary solution $u(\boldsymbol{x}, t) = u_0$ where u_0 is a real constant. We can set $u_0 = 0$ using the transformation $u \to u - u_0$. In the convection example, where u represents the vertical velocity of the fluid, the trivial solution, $u = 0$, would be the conduction state, where the fluid is at rest and there is no convection going on. We now assume that the trivial solution, $u = 0$, undergoes a stationary symmetry-breaking bifurcation at $\mu = 0$, being stable for $\mu < 0$ and unstable for $\mu > 0$. The bifurcation parameter μ is an external forcing that can be varied by the experimenter. As μ is increased through $\mu = 0$ the forcing becomes large enough to destabilise the uniform solution and any small perturbations or fluctuations resulting from noise in the experiment start to grow. In convection, μ would typically represent the temperature difference across the layer, with the origin shifted so that the critical temperature difference at which convection begins is set to $\mu = 0$. The system can now evolve in many different ways. There is no guarantee that it will

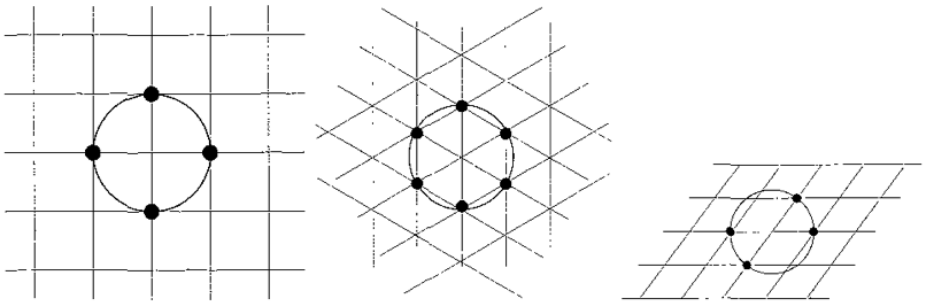


Fig. 5.3 Critical modes on doubly periodic dual lattices: (a) square lattice, (b) hexagonal lattice and (c) rhombic lattice

ever reach a stable end state, or that coherent spatial structures will develop, but we will focus on one possibility: the emergence of a lattice pattern.

Consider Fourier mode perturbations $e^{ik\cdot x}$ to $u = 0$. We assume that at the bifurcation point, $\mu = 0$, the zero solution is neutrally stable to a circle of modes with $|k| = k_c \neq 0$ and stable to all other modes, so that if a lattice pattern is selected from among that circle of modes it will be exactly periodic in space with a well-defined period at onset. For simplicity we assume that there is only one neutral circle.

At this point, what we want to do is to apply the centre manifold theorem to reduce the problem to something more tractable. Unfortunately we have two problems: first there is an infinite number of neutral modes lying on the circle $|k| = k_c$, so even if we could apply the theorem we would still have an infinite number of modes to deal with, and second there are decaying modes $|k| \neq k_c$ as close to the unit circle as you like, with growth rates as close to zero as you like, so the set of decaying modes does not have growth rates bounded away from zero. As things stand, equation (2.30) is not satisfied and we cannot apply the centre manifold theorem. Happily, if we insist that the solutions are doubly periodic with respect to a planar lattice, then all wavevectors k contributing to the pattern must lie at the vertices of a dual lattice, and our troubles are immediately resolved. As we can see from Figure 5.3, the critical circle intersects a finite number of vertices of the dual lattice, and all other lattice points are a finite distance away from it, ensuring that the growth rates of the corresponding modes are negative and bounded away from zero. It is worth pointing out now that in insisting the solutions be periodic on a lattice, we have made two choices: the choice to restrict our analysis to lattice patterns, and the choice of lattice – square over hexagonal, for example. These are quite reasonable choices: we know from observation that lattice patterns are

common, we have decided to carry out our analysis in a framework that accommodates them and we know that square lattice patterns are seen, so if we want to study them it is sensible to choose a square lattice. However they are restrictive choices: using this formalism we can't investigate why a system 'chooses' a lattice pattern over some other less regular structure, and in fact we can't even work out whether squares are preferred over hexagons, for example, because they don't fit on the same lattice. Nonetheless, the restriction to a lattice is a good start in the analysis of spatially periodic patterns, so let's see where we get with it.

We can now apply the centre manifold theorem, because the decaying modes all have negative growth rate bounded away from zero. This projects the dynamics onto a finite number of critical (neutral) modes, namely the equilibrium solutions $e^{i\boldsymbol{k}_j \cdot \boldsymbol{x}}$ of the linearised problem at the bifurcation point. We can now write

$$u(\boldsymbol{x}, t) = \sum_{j=1}^{n} z_j(t) e^{i\boldsymbol{k}_j \cdot \boldsymbol{x}} + c.c. \tag{5.7}$$

at leading order, where $z_j(t) \in \mathbb{C}$, $|\boldsymbol{k}_j| = k_c$. Note that n is half the number of modes lying on dual lattice vertices that are intersected by the neutral circle – we only need half the modes, because the amplitude of the mode $e^{-i\boldsymbol{k}_j \cdot \boldsymbol{x}}$ must be $\bar{z}_j$ since $u(\boldsymbol{x}, t)$ is real. As the bifurcation parameter μ increases past zero, the growth rate of the critical modes will become positive, and the amplitudes $z_j(t)$ will start to evolve according to a system of ordinary differential equations

$$\frac{\mathrm{d}z}{\mathrm{d}t} = g(z, \mu), \tag{5.8}$$

where $z = (z_1, \ldots, z_n) \in \mathbb{C}^n$. These are known as **amplitude equations**. The zero solution is now $z = \mathbf{0}$, and the requirement that it be stationary gives $g(\mathbf{0}, \mu) = \mathbf{0}$. If we temporarily consider equation (5.8) to be real, with $z = (\xi_1, \eta_1, \ldots, \xi_n, \eta_n) \in \mathbb{R}^{2n}$, where $z_j = \xi_j + i\eta_j$ for $j = 1, \ldots, n$, then the Jacobian is zero at the bifurcation point, $Dg|_{(\mathbf{0},0)} = 0$, since $Dg|_{(\mathbf{0},\mu)} = c(\mu)I$ with $c(0) = 0$, for a generic steady bifurcation at $\mu = 0$. Beyond the bifurcation point other modes on the lattice, higher harmonics of the critical modes, will also become unstable and contribute to $u(\boldsymbol{x}, t)$, so we should write

$$u(\boldsymbol{x}, t) = \sum_{j=1}^{n} z_j(t) e^{i\boldsymbol{k}_j \cdot \boldsymbol{x}} + c.c. + h.o.t., \tag{5.9}$$

where $h.o.t.$ denotes these higher order terms.

The ordinary differential equations (5.8) inherit some symmetry from the Euclidean symmetry of the original partial differential equation. The symmetries that are retained are those in the finite group of rotations and reflections that

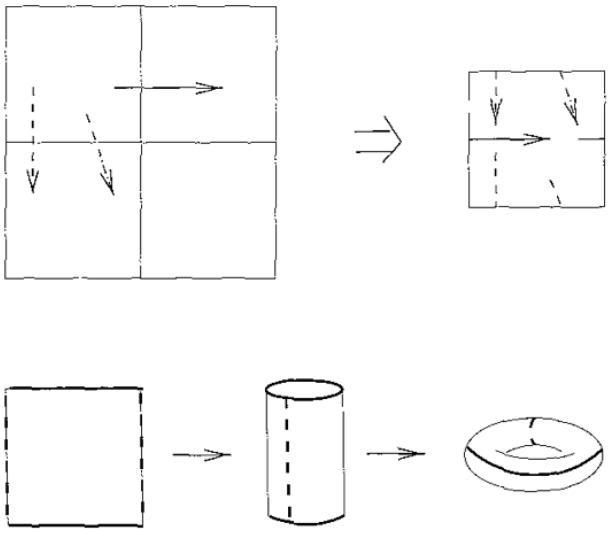


Fig 5 4. Translations on the square lattice form a torus

preserve the lattice, known as the **holohedry**, H, of the lattice, together with translations 'factored out by the lattice'. We are assuming periodicity on the lattice, so we can generate the full pattern from the solution in one cell of the lattice. If we have a translation out of the unit cell and into a neighbouring cell, then because the two cells are identical we can consider the translation to be out of the unit cell and then back in again on the other side. This is illustrated in Figure 5 4 for the square lattice, but the same thing holds for hexagonal or rhombic lattices. The translations effectively glue the opposite sides of the unit cell together to make a torus: this is factoring-out the translations by the lattice. We call the group of translations on the lattice, $T^2 \cong \mathbb{R}^2/\mathcal{L}$, the torus of translations, where here $\mathcal{L}$ is the group of translations that preserve the lattice and T^2 is a quotient group. So the symmetry group Γ of the bifurcation problem (5.8) is given by $\Gamma = H \ltimes T^2$. Another way of explaining the need to restrict the problem to a lattice is to say that the infinite plane is not a compact domain, but by restricting to a lattice we end up with a torus, which is compact. You might remember from Chapters 3 and 4 that many of the results we use to analyse bifurcations with symmetry depend on compactness. For the square and hexagonal lattices, we have $H = \mathcal{D}_4$ and $H = \mathcal{D}_6$ respectively. Equations (5 8) satisfy the equivariance condition

$$\gamma g(z, \mu) = g(\gamma z, \mu), \quad \forall \gamma \in \Gamma, \tag{5.10}$$

which allows us to write down the general form of the amplitude equations (5.8), in other words the normal form of the bifurcation. To see how this works, we

will work through the calculations for the two most common lattices: square and hexagonal.

5.3 Steady bifurcation on a square lattice

Consider a stationary bifurcation problem with Euclidean symmetry, restricted to a square lattice. The relevant symmetry group is $\mathcal{D}_4 \ltimes T^2$. By rescaling space appropriately we can set $k_c = 1$ to make the calculations look neater. Often when people rescale their equations it looks as if they simply invoke the magic word 'rescaling' and set various parameters to convenient values, but it is more legitimate than it might at first appear. To do the rescaling, set $x_j = \widetilde{x}_j / k_c$ for $j = 1, 2$, where $\boldsymbol{x} = (x_1, x_2)$, so that $e^{ik_c x_j} = e^{i\widetilde{x}_j}$, and then work in the new variables $\widetilde{x}_j$. In practice we immediately drop the tildes over the x_j, so it looks as if we have simply set $k_c = 1$. We will use the fundamental representation of $\mathcal{D}_4 \ltimes T^2$ on the square lattice in which there are two orthogonal critical wavevectors $(1, 0)$ and $(0, 1)$ and to leading order a real scalar solution takes the form

$$u(\boldsymbol{x}, t) = z_1(t)e^{ix_1} + z_2(t)e^{ix_2} + c.c. \quad , \quad z_1, z_2 \in \mathbb{C}. \tag{5.11}$$

The fundamental representation gives an accurate description of most experimentally observed lattice patterns, but in certain scenarios other choices of representation are required – this is discussed further in Section 6.1. We want to consider the action of $\mathcal{D}_4 \ltimes T^2$ on $\mathbb{C}^2$ ($\cong \mathbb{R}^4$). The group is generated by:

(i) a rotation through $\pi/2$, ρ, with $M_\rho = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ for the action on $\boldsymbol{x}$;

(ii) a reflection in x_1, m, with $M_m = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ for the action on $\boldsymbol{x}$;

(iii) translations $\boldsymbol{p} \in T^2$ whose action is given by $\boldsymbol{p} : \boldsymbol{x} \to \boldsymbol{x} + \boldsymbol{p}$

The actions of the generators on the amplitudes z_1 and z_2 can be deduced from the requirement that the general form of the solution for $u(\boldsymbol{x}, t)$ given in equation (5.11) should be unchanged under the symmetries.

The natural or **scalar** action of the Euclidean group on a function $v(\boldsymbol{x})$ is given by

$$\gamma v(\boldsymbol{x}) = v(\gamma^{-1}\boldsymbol{x}), \quad \forall \gamma \in E(2), \forall \boldsymbol{x} \in \mathbb{R}^2. \tag{5.12}$$

The inverse is needed so that the action of the group on the space of functions is a homomorphism. Writing the action of a group element γ more formally as $\theta(\gamma)$, where θ is a homomorphism, we need to satisfy

$$\theta(\gamma\sigma) = \theta(\gamma)\theta(\sigma), \quad \forall \sigma, \gamma \in E(2). \tag{5.13}$$

Now the action $\theta(\gamma)$ on the space of functions $f(\mathbb{R}^2)$, of which v is a member, is defined by

$$\theta(\gamma)v(\boldsymbol{x}) = v(\gamma^{-1}\boldsymbol{x}), \quad \forall \gamma \in E(2),\ \forall v \in f(\mathbb{R}^2),\ \forall \boldsymbol{x} \in \mathbb{R}^2, \tag{5.14}$$

which is just a more formal way of writing equation (5.12). If we define a function $w = \theta(\sigma)v \in f(\mathbb{R}^2)$, then we have

$$\begin{aligned}\theta(\gamma)(\theta(\sigma)v)(\boldsymbol{x}) &= \theta(\gamma)w(\boldsymbol{x}) = w(\gamma^{-1}\boldsymbol{x}) = v(\sigma^{-1}\gamma^{-1}\boldsymbol{x}) = v((\gamma\sigma)^{-1}\boldsymbol{x}) \\ &= \theta(\gamma\sigma)v(\boldsymbol{x}), \forall \sigma, \gamma \in E(2),\end{aligned} \tag{5.15}$$

since $w(\boldsymbol{x}) = \theta(\sigma)v(\boldsymbol{x}) = v(\sigma^{-1}\boldsymbol{x})$. The inverses ensure that the group elements act in the correct order on the righthand side.

So, in the example of $\mathcal{D}_4 \ltimes I^2$ the rotation acts according to $\rho u(\boldsymbol{x}, t) = u(\rho^{-1}\boldsymbol{x}, t)$. In this representation we have

$$M_{\rho^{-1}}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ -x_1 \end{pmatrix}, \tag{5.16}$$

and so applying the rotation, ρ, to to equation (5.11) gives

$$\rho u(\boldsymbol{x}, t) = z_1(t)e^{ix_2} + z_2(t)e^{-ix_1} + \bar{z}_1(t)e^{-ix_2} + \bar{z}_2(t)e^{ix_1}. \tag{5.17}$$

The general form of the solution on the righthand side has remained unchanged, but we have sent $(z_1, z_2) \to (\bar{z}_2, z_1)$. Thus we have $\rho : (z_1, z_2) \to (\bar{z}_2, z_1)$. Similarly we find the actions $m : (z_1, z_2) \to (\bar{z}_1, z_2)$ and $\boldsymbol{p} : (z_1, z_2) \to (e^{-ip_1}z_1, e^{-ip_2}z_2)$, where $\boldsymbol{p} = (p_1, p_2)$. This information will be useful when we come to derive the amplitude equations for z_1 and z_2. Considered as a real four-dimensional representation, the actions correspond to the following matrices

$$M_\rho = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad M_m = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{5.18}$$

$$M_{\boldsymbol{p}} = \begin{pmatrix} \cos p_1 & \sin p_1 & 0 & 0 \\ -\sin p_1 & \cos p_1 & 0 & 0 \\ 0 & 0 & \cos p_2 & \sin p_2 \\ 0 & 0 & -\sin p_2 & \cos p_2 \end{pmatrix}, \tag{5.19}$$

that act on the real vector, $\boldsymbol{z} = (\xi_1, \eta_1, \xi_2, \eta_2)$, where $z_1 = \xi + i\eta_1$ and $z_2 = \xi_2 + i\eta_2$. It is straightforward to check that the only matrices that commute with all three matrices are multiples of the identity, and so this is an absolutely irreducible representation.

Table 5.1 *The isotropy subgroups (up to conjugacy) for the steady bifurcation on a square lattice, using the fundamental representation, where* $\xi_1 = Re(z_1)$ *and* $\xi_2 = Re(z_2)$

Branch	Orbit representative (z_1, z_2)	Isotropy subgroup Σ (generators)	dim Fix(Σ)
Trivial solution	$(0, 0)$	$D_4 \ltimes T^2$	0
Squares	(ξ_1, ξ_1),	D_4 (ρ, m)	1
Rolls	$(\xi_1, 0)$,	$D_2 \ltimes S^1$ $(\rho^2, m, (0, p_2))$	1
Bimodal	(ξ_1, ξ_2), $\xi_1 \neq \xi_2$	D_2 (m, ρ^2)	2

The isotropy subgroups and their fixed-point subspaces are given in Table 5.1. The dimension of the fixed-point subspace is the same as the number of independent real amplitudes that we can vary in the orbit representative. We see that the two branches with one-dimensional fixed-point subspace are rolls and squares. The equivariant branching lemma guarantees that they will be primary branches. The isotropy subgroup of rolls, $D_2 \ltimes S^1$ is often given as $O(2) \times \mathbb{Z}_2$. Writing $D_2 \ltimes S^1$ emphasizes that rolls have rectangular symmetry plus translations in one direction.

You might have imagined that there would be further isotropy subgroups, such as $\mathbb{Z}_2 : (z_1, z_2) \to (\bar{z}_1, z_2)$, with orbit representative $(\xi_1, \xi_2 + i\eta_2)$, where $\eta_2 = Im(z_2)$, but in fact a point of this form has a second symmetry, a combination of a reflection and a translation in the x_2 direction, and is conjugate (by the translation) to the bimodal solution. Similarly, any other possibilities are conjugate by a translation in T^2 to one of the solutions in Table 5.1. This depends upon there being a lattice, so that we can have translations.

We shall now use the equivariance condition given by equation (5.10) to deduce the amplitude equations governing $z_1(t)$ and $z_2(t)$. To linear order, the equations can be expanded in the following form:

$$\frac{dz_1}{dt} = g_1(z_1, z_2) \equiv \mu_1 z_1 + c_1 \bar{z}_1 + c_2 z_2 + c_3 \bar{z}_2, \tag{5.20}$$

$$\frac{dz_2}{dt} = g_2(z_1, z_2) \equiv \mu_2 z_2 + d_1 \bar{z}_2 + d_2 z_1 + d_3 \bar{z}_1, \tag{5.21}$$

where the μ_j are bifurcation parameters, and the c_j and d_j are constants. In fact it is pretty clear that we must have $\mu_1 = \mu_2 \equiv \mu$ and $c_1 = c_2 = c_3 = d_1 = d_2 =$

$d_3 = 0$ or there wouldn't be a simple steady bifurcation at $\mu = 0$, but forget this for a minute while we take a look at the equivariance conditions

If we insist that equations (5.20) and (5.21) be equivariant under translations $p : (z_1, z_2) \to (e^{-ip_1}z_1, e^{-ip_2}z_2)$, then we must have

$$e^{-ip_1}(\mu_1 z_1 + c_1\bar{z}_1 + c_2 z_2 + c_3\bar{z}_2) = (\mu_1 e^{-ip_1}z_1 + c_1 e^{ip_1}\bar{z}_1 + c_2 e^{-ip_2}z_2 + c_3 e^{ip_2}\bar{z}_2), \tag{5.22}$$

$$e^{-ip_2}(\mu_2 z_2 + d_1\bar{z}_2 + d_2 z_1 + d_3\bar{z}_1) = (\mu_2 e^{-ip_2}z_2 + d_1 e^{ip_2}\bar{z}_2 + d_2 e^{-ip_1}z_1 + d_3 e^{ip_1}\bar{z}_1) \tag{5.23}$$

for all p_1, p_2, z_1 and z_2, and so the coefficients c_j and d_j must be zero for $j = 1, 2, 3$

Applying rotation equivariance $\rho : (z_1, z_2) \to (\bar{z}_2, z_1)$ in a similar way leads to the equations

$$\bar{\mu}_2\bar{z}_2 = \mu_1\bar{z}_2, \tag{5.24}$$

$$\mu_1 z_1 = \mu_2 z_1, \tag{5.25}$$

which tell us that $\mu_1 = \mu_2 \equiv \mu$, where μ is real. Thus, just as we suspected, our linear equations are

$$\frac{dz_1}{dt} = \mu z_1, \tag{5.26}$$

$$\frac{dz_2}{dt} = \mu z_2. \tag{5.27}$$

The trick is to apply this procedure to a higher order expansion of the amplitude equations. The equivariance condition (5.10) implies that we can do this order by order. Usually expanding to cubic order is enough, but even then the most general form of the amplitude equations would be very messy to write down. Luckily the requirement of equivariance under translations eliminates most of the terms straightaway. For example the possible quadratic terms that could appear in the dz_1/dt equation, together with their transformations under a general translation (p_1, p_2), are

$$\begin{aligned} z_1^2 &\to z_1^2 e^{-2ip_1}, \quad \bar{z}_1^2 \to \bar{z}_1^2 e^{2ip_1}, \\ z_2^2 &\to z_2^2 e^{-2ip_2}, \quad \bar{z}_2^2 \to \bar{z}_2^2 e^{2ip_2}, \\ z_1 z_2 &\to z_1 z_2 e^{-i(p_1+p_2)}, \quad \bar{z}_1\bar{z}_2 \to \bar{z}_1\bar{z}_2 e^{i(p_1+p_2)}, \\ z_1\bar{z}_2 &\to z_1\bar{z}_2 e^{-i(p_1-p_2)}, \quad \bar{z}_1 z_2 \to \bar{z}_1 z_2 e^{i(p_1-p_2)}. \end{aligned} \tag{5.28}$$

We know that the lefthand side of the equation for dz_1/dt is multiplied by e^{-ip_1} under a general translation. None of the quadratic terms transforms in the same

way, so none can appear on the righthand side. It is straightforward to check that of all the possible cubic terms only $|z_1|^2 z_1$ ($\rightarrow |z_1|^2 z_1 e^{-ip_1}$) and $|z_2|^2 z_1$ ($\rightarrow |z_2|^2 z_1 e^{-ip_1}$) transform in the appropriate way, and so these are the only cubic terms that can be included on the righthand side of the dz_1/dt equation. Similarly the nonlinear terms that can appear in the dz_2/dt equation can be deduced from the requirement that they be multiplied by e^{-ip_2} under a general translation. After a little thought it is easy to see that to cubic order the amplitude equations must take the form

$$\frac{dz_1}{dt} = \mu z_1 - a_1|z_1|^2 z_1 - a_2|z_2|^2 z_1, \tag{5.29}$$

$$\frac{dz_2}{dt} = \mu z_2 - b_1|z_1|^2 z_2 - b_2|z_2|^2 z_2, \tag{5.30}$$

where a_1, a_2, b_1 and b_2 are constants

Now using reflection equivariance $m : (z_1, z_2) \rightarrow (\bar{z}_1, z_2)$ gives

$$\begin{aligned}\frac{d\bar{z}_1}{dt} &= \mu\bar{z}_1 - a_1|z_1|^2\bar{z}_1 - a_2|z_2|^2\bar{z}_1 \\ &= \overline{\mu z_1 - a_1|z_1|^2 z_1 - a_2|z_2|^2 z_1} = \overline{\frac{dz_1}{dt}},\end{aligned} \tag{5.31}$$

$$\frac{dz_2}{dt} = \mu z_2 - b_1|z_1|^2 z_2 - b_2|z_2|^2 z_2, \tag{5.32}$$

which leads to the requirement that a_1 and a_2 be real

Finally rotation equivariance $\rho : (z_1, z_2) \rightarrow (\bar{z}_2, z_1)$ gives the equations

$$\frac{d\bar{z}_2}{dt} = \mu\bar{z}_2 - a_1|z_2|^2\bar{z}_2 - a_2|z_1|^2\bar{z}_2, \tag{5.33}$$

$$\frac{dz_1}{dt} = \mu z_1 - b_1|z_2|^2 z_1 - b_2|z_1|^2 z_1, \tag{5.34}$$

which agree with the originals (5.29) and (5.30) as long as $b_1 = a_2$ and $b_2 = a_1$

So in the end, the $\mathcal{D}_4 \ltimes I^2$-equivariant amplitude equations truncated at cubic order turn out to be

$$\frac{dz_1}{dt} = \mu z_1 - a_1|z_1|^2 z_1 - a_2|z_2|^2 z_1, \tag{5.35}$$

$$\frac{dz_2}{dt} = \mu z_2 - a_1|z_2|^2 z_2 - a_2|z_1|^2 z_2, \tag{5.36}$$

where μ, a_1 and a_2 are real.

We can use these equations to look at the existence and stability of periodic patterns on a square lattice. First we set $dz_1/dt = dz_2/dt = 0$ and find the following stationary solutions:

(i) the trivial solution $z_1 = z_2 = 0$;
(ii) x_1-rolls $|z_1|^2 = \mu/a_1$, $z_2 = 0$;
(iii) x_2-rolls $z_1 = 0$, $|z_2|^2 = \mu/a_1$;
(iv) squares $|z_1|^2 = |z_2|^2 = \mu/(a_1 + a_2)$

It is always possible to set the imaginary parts of any solution (z_1, z_2) to zero, by moving the origin. If, for example, we have x_1-rolls with $z_1 = \sqrt{\mu/a_1}e^{ix_0}$, where x_0 is some real constant, then reconstructing the marker quantity $u(\boldsymbol{x}, t)$ according to

$$u(\boldsymbol{x}, t) = z_1(t)e^{ix_1} + z_2(t)e^{ix_2} + c.c. + h.o.t. \tag{5.37}$$

gives

$$u(\boldsymbol{x}, t) = \sqrt{\mu/a_1}e^{ix_0}e^{ix_1} + c.c. + h.o.t. \tag{5.38}$$

Clearly if we move the origin by a distance x_0 in the negative x_1 direction, so that $x_1 \to \tilde{x}_1 - x_0$ where $\tilde{x}_1$ is the new x_1 variable, we recover

$$u(\boldsymbol{x}, t) = \sqrt{\mu/a_1}e^{i\tilde{x}_1} + c.c. + h.o.t\ , \tag{5.39}$$

and $z_1 = \sqrt{\mu/a_1}$ in the new variables with the shifted origin. Since by construction the amplitude equations are equivariant with respect to translations, they are unchanged by this manoeuvre. If the solution (z_1, z_2) is real, then the righthand sides of equations (5.35) and (5.36) are real, so if (z_1, z_2) starts out real then the imaginary parts never grow, and (z_1, z_2) stays real for all time. This is very handy because it means that we can set (z_1, z_2) real initially by shifting the origin appropriately, and then assume that it is real ever after.

The zero growth rates of the imaginary parts are a consequence of equivariance under translations in the x_1 and x_2 directions. We saw in Chapter 3 that a continuous symmetry leads to a zero eigenvalue with eigenvector tangent to the group orbit. If we take any real solution (r_1, r_2), $r_1, r_2 \in \mathbb{R}$, and apply a translation $(p_1, 0)$, we get $(r_1e^{-ip_1}, r_2)$. Differentiating this with respect to p_1 and evaluating the result at $p_1 = 0$ gives an eigenvector $(-ir_1, 0)$, $r_1 \in \mathbb{R}$, in the imaginary z_1 direction, that has zero growth rate eigenvalue according to equation (4.13). We can do the same thing for translations in the x_2 direction to find an eigenvector in the imaginary z_2 direction with zero growth rate eigenvalue. Another way of saying all this is that the zero growth rates are related to the freedom to choose the origins of x_1 and x_2. The system has no preference for any one position over another, so it is linearly neutrally stable to transformations that shift the origin.

If we choose (z_1, z_2) to be real, the amplitude equations become

$$\frac{dz_1}{dt} = \mu z_1 - a_1 z_1^3 - a_2 z_2^2 z_1, \tag{5.40}$$

$$\frac{dz_2}{dt} = \mu z_2 - a_1 z_2^3 - a_2 z_1^2 z_2, \tag{5.41}$$

where all the variables and constants are real. These equations are just a restatement of equation (4.72) that we found for the bifurcation with $\mathcal{D}_4$ symmetry under the natural representation, because by setting the imaginary parts of the amplitudes to zero we have effectively 'removed' the translations on the lattice.

Exactly the same stability analysis that we used in that case applies here, because the equations are the same. The possible bifurcation diagrams can be read from Figure 4.10, where now the labels R and S refer to rolls and squares respectively.

Notice that the bimodal solution does not turn up as a solution of the amplitude equations truncated at cubic order. The truncation loses some information, and if we wanted to find the bimodal solution we would have to expand the amplitude equations to higher order, in fact fifth order in this case. So to summarise, considering the isotropy subgroups will tell you all the possible solutions categorised according to their symmetries, but gives limited information about their stability, whereas the amplitude equations will give you detailed information on the shape of the branches and their stability, but may miss some solutions if you don't go to high enough order.

Having looked at the square lattice, we will now move on and consider a lattice of hexagons.

5.4 Steady bifurcation on a hexagonal lattice

Once again we are considering the stationary bifurcation with Euclidean symmetry, this time restricted to a hexagonal lattice, so the symmetry group is $\mathcal{D}_6 \ltimes T^2$. We use the fundamental representation, writing a scalar solution to leading order as

$$u(\boldsymbol{x}, t) = \sum_{j=1}^{3} z_j(t) e^{i\boldsymbol{k}_j \cdot \boldsymbol{x}} + c.c., \quad z_j \in \mathbb{C} \tag{5.42}$$

and we consider the action of $\mathcal{D}_6 \ltimes T^2$ on $\mathbb{C}^3$. For a hexagonal lattice, the wavevectors are at angles of $2\pi/3$ to each other and must therefore satisfy $\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3 = \boldsymbol{0}$ (see Figure 5.5). We call this relationship between the wavevectors a **resonance**. Resonances occur when any one wavevector that appears in a representation can be written as a linear combination of the others.

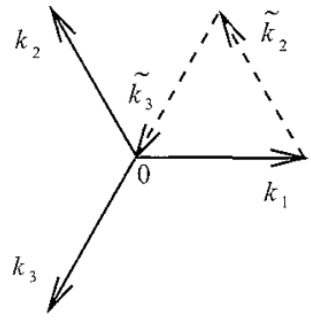


Fig 5 5 The arrangement of the basic wavevectors for the hexagonal lattice The wavevectors $\tilde{k_2}$ and $\tilde{k_3}$ are shifted copies of k_2 and k_3, showing that $k_1 + k_2 + k_3 = 0$

We will use the following symmetries to characterise the solutions:

(i) a rotation through $2\pi/3$, $\rho : (z_1, z_2, z_3) \to (z_3, z_1, z_2)$;
(ii) a reflection through the origin (or rotation through π), $x \to -x$, $m_0 : (z_1, z_2, z_3) \to (\bar{z}_1, \bar{z}_2, \bar{z}_3)$;
(iii) a reflection in a vertical plane $m_\nu : (z_1, z_2, z_3) \to (z_1, z_3, z_2)$;
(iv) translations $p \in I^2$ whose action is given by $p \cdot (z_1, z_2, z_3) = (e^{-ik_1 p} z_1, e^{-ik_2 p} z_2, e^{-ik_3 p} z_3)$

In fact the first two could be replaced by a single generator, rotation through $\pi/3$, but this symmetry doesn't take quite such a nice form, so we shall work with the set above to ensure neatness in the calculations In working out the amplitude equations later on it will be important to note that $\bar{z}_2\bar{z}_3$ transforms under translations p to $\bar{z}_2\bar{z}_3 e^{i(k_2+k_3)p} = \bar{z}_2\bar{z}_3 e^{-ik_1 p}$ So $\bar{z}_2\bar{z}_3$ transforms in the same way as z_1 and can appear in the equation for dz_1/dt This also means that we can't play our trick of making all the amplitudes real by shifting the origin, because if we start with z_1, z_2 and z_3 all complex and shift the origin by p so that z_1 and z_2, for example, are real, then the phase of z_3 will have changed by $e^{i(k_1+k_2)p}$, which will not in general be the right amount of phase shift to make z_3 real.

Considered as a representation over $\mathbb{R}^6$ the action of the group is, once again, absolutely irreducible, so we can use the equivariant branching lemma The isotropy subgroups of $D_6 \ltimes I^2$, with fixed-point subspaces of dimension no greater than two are given in Table 5 2 We see that there are two isotropy subgroups with one-dimensional fixed-point subspace, $D_2 \ltimes S^1 (\cong O(2) \times \mathbb{Z}_2)$ corresponding to rolls and D_6 corresponding to hexagons So according to the equivariant branching lemma there must be two primary branches at the bifurcation

Table 5.2 *The isotropy subgroups with fixed-point subspace of dimension no greater than two, for the steady bifurcation on a hexagonal lattice, using the fundamental representation, where* $z_j = \xi_j + i\eta_j$ *for* $j = 1, 2, 3$. *After Golubitsky, Swift and Knobloch (1984)*

Branch	Orbit representative (z_1, z_2, z_3)	Isotropy subgroup Σ (generators)	dim Fix(Σ)
Trivial solution	$(0, 0, 0)$	$D_6 \ltimes T^2$	0
Rolls	$(\xi_1, 0, 0)$	$D_2 \ltimes S^1$ $(m_v, m_o, (0, p_2) \in T^2)$	1
Hexagons	(ξ_1, ξ_1, ξ_1)	D_6 $(m_v, m_o\rho)$	1
Up-hexagons	$\xi_1 > 0$		
Down-hexagons	$\xi_1 < 0$		
Rectangles	(ξ_1, ξ_2, ξ_2)	D_2 (m_v, m_o)	2
Up-rectangles	$\xi_1 > 0$		
Down-rectangles	$\xi_1 < 0$		
Patchwork quilt	$\xi_1 = 0$		
Triangles	$(\xi_1 + i\eta_1, \xi_1 + i\eta_1, \xi_1 + i\eta_1)$	D_3 (m_v, ρ)	2
Regular triangles	$\xi_1 = 0$		

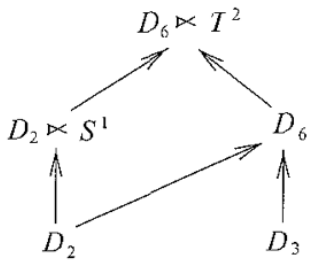


Fig. 5.6 Part of the isotropy lattice for the bifurcation on the hexagonal lattice using the fundamental representation of the lattice. Inclusion is indicated by an arrow.

point, one of rolls and one of hexagons. The top part of the isotropy lattice is shown in Figure 5.6. The solutions described in Table 5.2 are illustrated in Figure 5.7.

Now we want to use equivariance to tell us what the amplitude equations should be. As we saw earlier, the combination $\bar{z}_2\bar{z}_3$ will transform like z_1 under translations. So using translation equivariance alone, it is easy to conclude that up to cubic

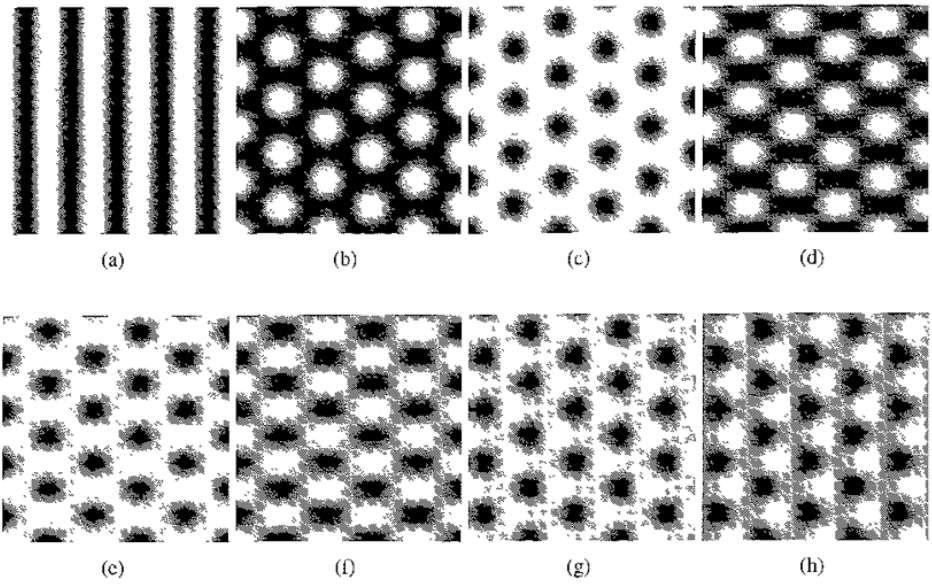


Fig 5 7 Greyscale plots of some solutions on the hexagonal lattice: (a) rolls, (b) up hexagons, (c) down hexagons, (d) up rectangles, (e) down rectangles, (f) patchwork quilt, (g) triangles and (h) regular triangles

order the amplitude equations ought to look something like this:

$$\frac{dz_1}{dt} = \mu_1 z_1 + a_1 \bar{z}_2 \bar{z}_3 - b_1 |z_1|^2 z_1 - c_1 |z_2|^2 z_1 - d_1 |z_3|^2 z_1, \tag{5.43}$$

$$\frac{dz_2}{dt} = \mu_2 z_2 + a_2 \bar{z}_3 \bar{z}_1 - b_2 |z_2|^2 z_2 - c_2 |z_3|^2 z_2 - d_2 |z_1|^2 z_2, \tag{5.44}$$

$$\frac{dz_3}{dt} = \mu_3 z_3 + a_3 \bar{z}_1 \bar{z}_2 - b_3 |z_3|^2 z_3 - c_3 |z_1|^2 z_3 - d_3 |z_2|^2 z_3, \tag{5.45}$$

where the μ_i, for $i = 1, 2, 3$, are bifurcation parameters, and the a_i, b_i, c_i and d_i, for $i = 1, 2, 3$, are complex constants Applying equivariance under the rotation $\rho : (z_1, z_2, z_3) \to (z_3, z_1, z_2)$ we see that we must have $\mu_1 = \mu_2 = \mu_3 \equiv \mu$, $a_1 = a_2 = a_3 \equiv a$, $b_1 = b_2 = b_3 \equiv b$, $c_1 = c_2 = c_3 \equiv c$ and $d_1 = d_2 = d_3 \equiv d$

Equivariance under reflection through the origin $m_0 : (z_1, z_2, z_3) \to (\bar{z}_1, \bar{z}_2, \bar{z}_3)$ shows that μ, a, b, c and d must all be real

Equivariance under reflection in the vertical plane $m_v : (z_1, z_2, z_3) \to (z_1, z_3, z_2)$ leads to the requirement that $d = c$

The final equations to cubic order then are

$$\frac{dz_1}{dt} = \mu z_1 + a\bar{z}_2\bar{z}_3 - b|z_1|^2 z_1 - c(|z_2|^2 + |z_3|^2)z_1, \tag{5.46}$$

$$\frac{dz_2}{dt} = \mu z_2 + a\bar{z}_3\bar{z}_1 - b|z_2|^2 z_2 - c(|z_3|^2 + |z_1|^2)z_2, \tag{5.47}$$

$$\frac{dz_3}{dt} = \mu z_3 + a\bar{z}_1\bar{z}_2 - b|z_3|^2 z_3 - c(|z_1|^2 + |z_2|^2)z_3. \tag{5.48}$$

Before we use these to look at the existence and stability of solution branches, let's take a quick diversion to see how the most general form for the amplitude equations could be obtained The four polynomials

$$s_1 = u_1 + u_2 + u_3, \tag{5.49}$$

$$s_2 = u_1u_2 + u_2u_3 + u_3u_1, \tag{5.50}$$

$$s_3 = u_1u_2u_3, \tag{5.51}$$

$$q = z_1z_2z_3 + \bar{z}_1\bar{z}_2\bar{z}_3, \tag{5.52}$$

where $u_i = |z_i|^2$, are invariant under all elements of the group, and hence so is any real function of them $h(s_1, s_2, s_3, q)$. It can be shown (see Buzano & Golubitsky, 1983) that all smooth functions invariant under $\mathcal{D}_6 \ltimes T^2$ take this form, and that the equivariance condition (5 10) implies that the amplitude equation for z_1 must take the form

$$\frac{dz_1}{dt} = z_1\big(h_1 + u_1h_3 + u_1^2h_5\big) + \bar{z}_2\bar{z}_3\big(g_2 + u_1g_4 + u_1^2g_6\big), \tag{5.53}$$

where the h_j and g_j are functions of s_1, s_2, s_3, q and the bifurcation parameter μ Equations for dz_2/dt, dz_3/dt are obtained by cyclic permutation of $\{z_1, z_2, z_3\}$ and $\{u_1, u_2, u_3\}$ according to rotation equivariance Higher order terms are not required inside the brackets, since

$$u_1^3 = s_3 - s_2u_1 + s_1u_1^2. \tag{5.54}$$

Getting back to the cubic truncation of the amplitude equations, we can look directly at the existence and stability of the roll and hexagon solution branches We can choose $a > 0$ in equations (5.46)–(5 48) without loss of generality since the equations are invariant under the transformation $(a \to -a, z \to -z)$ So if the coefficient a were actually negative in a given situation, we could simply define a new set of amplitudes $\hat{z} = -z$ and they would evolve according to the same amplitude equations, but with $\hat{a} = -a > 0$. Henceforth we assume $a > 0$. The equilibrium solutions to the cubic truncation of the amplitude equations are

(i) the trivial solution $z_1 = z_2 = z_3 = 0$;
(ii) rolls, for example $z_1 = \sqrt{\mu/b}$, $z_2 = z_3 = 0$;
(iii) hexagons $Re(z_1) = Re(z_2) = Re(z_3) = R_0$, where

$$\mu + aR_0 - (b + 2c)R_0^2 = 0,$$
$$Im(z_1) = Im(z_2) = Im(z_3) = 0; \tag{5.55}$$

(iv) rectangles

$$Re(z_1) = -\frac{a}{(b-c)},$$
$$Re(z_2) = Re(z_3) = \pm\sqrt{\frac{1}{b+c}\left(\mu - \frac{a^2 b}{(b-c)^2}\right)},$$
$$Im(z_1) = Im(z_2) = Im(z_3) = 0 \tag{5.56}$$

Clearly not all of the solutions given in Table 5.2 are equilibrium solutions of the equations truncated at third order: some require higher-order terms. The roll and hexagon branches are primary, bifurcating from the zero solution at $\mu = 0$, just as predicted by the equivariant branching lemma, since both have one-dimensional fixed-point subspaces. However the rectangle branch is secondary; it connects the roll branch, at $\mu = a^2 b/(b-c)^2$, to the hexagon branch, at $\mu = (2b + c)a^2/(b-c)^2$, as shown in Figure 5.8. This is consistent with the isotropy subgroup corresponding to rectangles having a two-dimensional fixed-point subspace.

We can work out the stability of each of the equilibrium solutions using the cubic truncation of the amplitude equations. Clearly the zero solution is stable to small perturbations if $\mu < 0$ and unstable if $\mu > 0$, just as it is on a square lattice. Let us look at perturbations to rolls by setting $z_1 = R_0(1 + \delta z_1)$, $z_2 = \delta z_2$, $z_3 = \delta z_3$, where $R_0 = \sqrt{\mu/b}$ and where all the perturbations are complex with $|\delta z_1|, |\delta z_2|, |\delta z_3| \ll 1$. Linearising we find

$$\frac{\mathrm{d}}{\mathrm{d}t}(\delta z_1 + \overline{\delta z_1}) = -2\mu(\delta z_1 + \overline{\delta z_1}), \tag{5.57}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(\delta z_1 - \overline{\delta z_1}) = 0, \tag{5.58}$$

$$\frac{\mathrm{d}(\delta z_2)}{\mathrm{d}t} = \mu(1 - c/b)\delta z_2 + a\sqrt{\frac{\mu}{b}}\overline{\delta z_3}, \tag{5.59}$$

$$\frac{\mathrm{d}(\delta z_3)}{\mathrm{d}t} = \mu(1 - c/b)\delta z_3 + a\sqrt{\frac{\mu}{b}}\overline{\delta z_2} \tag{5.60}$$

The first two growth rate eigenvalues are clearly $\sigma_1 = -2\mu$ and $\sigma_2 = 0$. To find the remaining four from the last two complex equations, we set $\delta z_2 = \widehat{\delta z_2} e^{\sigma t}$ and

$\delta z_3 = \widehat{\delta z_3} e^{\sigma t}$, where $\widehat{\delta z_2}$ and $\widehat{\delta z_3}$ are arbitrary complex constants, and solve the resulting eigenvalue problem to find

$$\sigma_{3\,4\,5\,6} = \mu(b-c)/b \pm a\sqrt{\mu/b} \quad \text{(twice each)} \tag{5.61}$$

So rolls are (neutrally) linearly stable if $\mu > 0$, $(b-c) < 0$ and $\sqrt{\mu/b} > -a/(b-c)$

Perturbing a hexagonal solution so that $z_1 = R_0(1+\delta z_1)$, $z_2 = R_0(1+\delta z_2)$, $z_3 = R_0(1+\delta z_3)$, where R_0 satisfies equation (5.55) and $|\delta z_1|, |\delta z_2|, |\delta z_3| \ll 1$, and linearising gives

$$\begin{aligned}\frac{\mathrm{d}(\delta z_1)}{\mathrm{d}t} &= aR_0(-\delta z_1 + \overline{\delta z_2} + \overline{\delta z_3}) - bR_0^2(\delta z_1 + \overline{\delta z_1}) \\ &\quad -cR_0^2(\delta z_2 + \overline{\delta z_2} + \delta z_3 + \overline{\delta z_3}),\end{aligned} \tag{5.62}$$

together with similar equations for $\mathrm{d}(\delta z_2)/\mathrm{d}t$ and $\mathrm{d}(\delta z_3)/\mathrm{d}t$ obtained by cyclically permuting the indices $\{1, 2, 3\}$. From these we deduce

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}(\delta z_1 + \overline{\delta z_1} + \delta z_2 + \overline{\delta z_2} + \delta z_3 + \overline{\delta z_3}) \\ &= (aR_0 - 2(b+2c)R_0^2)(\delta z_1 + \overline{\delta z_1} + \delta z_2 + \overline{\delta z_2} + \delta z_3 + \overline{\delta z_3}),\end{aligned} \tag{5.63}$$

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}(\delta z_i + \overline{\delta z_i} - \delta z_j - \overline{\delta z_j}) \\ &(-2aR_0 - 2(b-c)R_0^2)(\delta z_i + \overline{\delta z_i} - \delta z_j - \overline{\delta z_j}), \quad i \neq j\end{aligned} \tag{5.64}$$

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}(\delta z_1 - \overline{\delta z_1} + \delta z_2 - \overline{\delta z_2} + \delta z_3 - \overline{\delta z_3}) \\ &= -3aR_0(\delta z_1 - \overline{\delta z_1} + \delta z_2 - \overline{\delta z_2} + \delta z_3 - \overline{\delta z_3}),\end{aligned} \tag{5.65}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(\delta z_i - \overline{\delta z_i} - \delta z_j + \overline{\delta z_j}) = 0, \quad i \neq j \tag{5.66}$$

and so the growth rate eigenvalues are

$$\sigma_1 = -3aR_0, \tag{5.67}$$

$$\sigma_{2\,3} = 0 \quad \text{(twice)}, \tag{5.68}$$

$$\sigma_4 = aR_0 - 2(b+2c)R_0^2, \tag{5.69}$$

$$\sigma_{5,6} = -2aR_0 - 2(b-c)R_0^2 \quad \text{(twice)}. \tag{5.70}$$

So hexagons are stable if $R_0 > 0$, $aR_0 - 2(b+2c)R_0^2 < 0$ and $-2aR_0 - 2(b-c)R_0^2 < 0$

Similarly rectangles can be shown always to be unstable. The bifurcation diagram is shown in Figure 5 8.

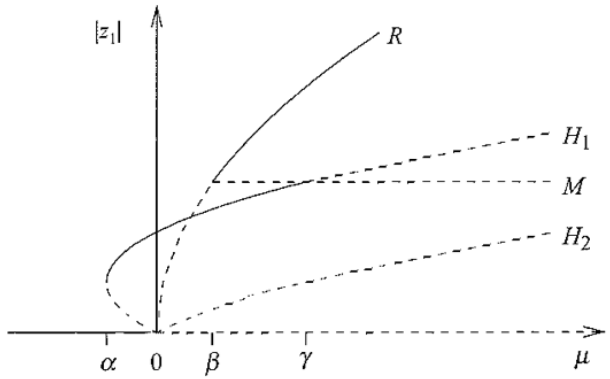


Fig. 5.8 Bifurcation diagram for hexagons in the case $(b - c) < 0$. The amplitude $|z_1|$ is plotted against μ for the roll-solution branch (labelled R) with $z_2 = z_3 = 0$, for the rectangle (mixed-mode) branch (labelled M) with $|z_1|$ constant and $|z_2| = |z_3|$, and for the two hexagon-solution branches with $|z_1| = |z_2| = |z_3|$. The curves labelled H_1 and H_2 correspond to the hexagon solutions with $\mathrm{Arg}(z_1 z_2 z_3) = 0$ (up-hexagons) and $\mathrm{Arg}(z_1 z_2 z_3) = \pi$ (down-hexagons) respectively. Branches where all three components of the solution are stable are indicated by solid curves, and branches where one or more components are unstable are indicated by dashed curves. Consequently, the behaviour of three growth rate eigenvalues is projected onto the bifurcation diagram for one component. Both up-hexagons and the trivial solution are stable in the range $\alpha \equiv -a^2/4(b + 2c) < \mu < 0$, and both rolls and up-hexagons are stable in the range $\beta \equiv a^2 b/(b - c)^2 < \mu < a^2(2b + c)/(b - c)^2 \equiv \gamma$.

There is hysteresis in the system: both hexagons and the trivial solution are stable in the range $-a^2/4(b + 2c) < \mu < 0$, and both rolls and hexagons are stable in the range $a^2 b/(b - c)^2 < \mu < a^2(2b + c)/(b - c)^2$, as shown in Figure 5.8. The lower branch of hexagons is unstable, because $aR_0 - 2(b + 2c)R_0^2 > 0$ holds there; the instability takes the form of a growing hexagonal amplitude disturbance, where $\delta z_1 + \overline{\delta z_1} + \delta z_2 + \overline{\delta z_2} + \delta z_3 + \overline{\delta z_3}$ grows, but $\delta z_i + \overline{\delta z_i} - \delta z_j - \overline{\delta z_j}$ for $i \neq j$ does not. The system is attracted to the stable upper solution branch. There is a saddle-node bifurcation at $R_0 = a/2(b + 2c)$ $(\mu = -a^2/4(b + 2c))$ at which the unstable hexagons gain stability, moving from the lower to the upper branch. There is also a bifurcation to rectangles at $R_0 = -a/(b - c)$ $(\mu = a^2(2b + c)/(b - c)^2)$, because $-2aR_0 - 2(b - c)R_0^2 > 0$ holds there. Here $\delta z_1 + \overline{\delta z_1} + \delta z_2 + \overline{\delta z_2} + \delta z_3 + \overline{\delta z_3}$ does not grow, but $\delta z_i + \overline{\delta z_i} - \delta z_j - \overline{\delta z_j}$ for $i \neq j$ does. This will happen at positive R_0 only if $b < c$, which is the situation that ensures that rolls are the stable solution for large μ, as is observed in practice (Busse, 1967). In fact, here we see the hysteretic transition to rolls. This was observed experimentally by Dubois, Bergé and Wesfreid (1978) during convection in water near 4 °C.

5.4.1 Additional $\mathbb{Z}_2$ symmetry (midplane reflection)

There is a symmetry between the regions of negative and positive u for rolls in Figure 5.1: they are the same size and shape and arranged periodically next to each other. On the other hand hexagons are asymmetric: the negative and positive regions are different in shape, with positive being in the middle of the hexagons and negative around the outside. This difference is associated with the presence or absence of an additional symmetry $m_h : z \to -z$, where $z = (z_1, z_2, z_3)$. In this section we shall examine what happens when this reflection symmetry is imposed on the bifurcation on the hexagonal lattice. Two things follow immediately: first the symmetry group of the bifurcation problem is now $(\mathcal{D}_6 \times \mathbb{Z}_2) \ltimes T^2$, and second no even terms are permitted in the amplitude equations. Now the reflection symmetry acts on $u(x, t)$ through the amplitudes z, and so we also have $u(x, t) \to -u(x, t)$. In convection $u(x, t)$ typically represents the vertical fluid velocity, so the reflection symmetry expresses the constraint that there should be symmetry between the up and down directions, in other words there is a reflection symmetry in the midplane of the convection layer. If the reflection symmetry is broken it implies that the up and down directions are not equivalent: there is some fundamental difference between the top of the convection layer and the bottom. One example of this is when convection takes place in an open container so that the fluid is open to the air at the top, but not at the bottom. Surface tension comes into play at the free top surface, but not at the bottom. This is known as Bénard–Marangoni convection: hexagons are favoured over rolls because hexagons display up-down asymmetry, while rolls do not. On the other hand in Rayleigh–Bénard convection the container is closed at the top and the convecting fluid fills the entire layer between the top and bottom of the container, so there is no surface tension acting and in the absence of other symmetry-breaking there is up-down symmetry and rolls are preferred. You might argue that there is always an obvious up-down asymmetry because after all you are heating the layer from below and not from above. This is quite true, but in formulating the bifurcation problem we are only interested in the perturbation from the asymmetric temperature profile of the conduction state, and this perturbation is zero at both the top and bottom of the layer: more detail can be found in Chapter 1. The up-down symmetry breaking explains the 'up-hexagons' and 'down-hexagons' nomenclature of Table 5.2: up-hexagons have fluid flowing up ($u > 0$) in the middle and down ($u < 0$) around the edges, while for down-hexagons it is the other way round. Systems that lack the $z \to -z$ symmetry will tend to prefer one of up-hexagons or down-hexagons over the other, whereas a symmetric system will not distinguish between them and initial conditions will determine which is seen

Table 5.3. *The axial patterns for the steady bifurcation on a hexagonal lattice with additional $\mathbb{Z}_2$ symmetry, using the fundamental representation of the lattice, and writing $z_j = \xi_j + i\eta_j$ for $j = 1, 2, 3$.*

The notation $\{\gamma, (p_1, p_2)\}$ denotes the composition of a reflection or rotation, γ, with a translation $(p_1, p_2) \in T^2$. The notation $\mathbb{Z}_2^-$ denotes a **twisted** version of $\mathbb{Z}_2$ that is generated by reflection composed with a translation. After Golubitsky, Swift and Knobloch (1984).

Branch	Orbit representative (z_1, z_2, z_3)	Isotropy subgroup Σ (generators)
Rolls	$(\xi_1, 0, 0)$	$(\mathcal{D}_2 \times \mathbb{Z}_2^-) \ltimes S^1$ $(m_v, m_o, (0, p_2), \{m_h, (\pi, \pi/\sqrt{3})\})$
Hexagons	(ξ_1, ξ_1, ξ_1)	$\mathcal{D}_6$ $(m_v, m_o\rho)$
Patchwork quilt	$(0, \xi_2, \xi_2)$	$\mathcal{D}_2 \times \mathbb{Z}_2^-$ $(m_v, m_o, \{m_h, (0, 2\pi/\sqrt{3})\})$
Regular triangles	$(i\eta_1, i\eta_1, i\eta_1)$	$\mathcal{D}_3 \times \mathbb{Z}_2^-$ $(m_v, \rho, m_o m_h)$

The only rectangular or triangular solutions permitted by the additional reflection symmetry are the patchwork quilt and regular triangles respectively. These solutions have only one independent amplitude. Since each one is now the most general solution having its given isotropy subgroup, the isotropy subgroup has one-dimensional fixed-point subspace and a branch of these solutions is guaranteed by the equivariant branching lemma alongside rolls and hexagons. The axial patterns, their isotropy subgroups and generators are given in Table 5.3.

To cubic order the amplitude equations can be derived from those for the case without midplane reflection (equations (5.46)–(5.48)), by setting $a = 0$. However it turns out that we need to go to quintic order to distinguish between hexagons and regular triangles, so the amplitude equations we need are

$$\begin{aligned}\frac{dz_1}{dt} &= \mu z_1 - b|z_1|^2 z_1 - c(|z_2|^2 + |z_3|^2)z_1 - d_1|z_1|^4 z_1 \\ &\quad -d_2(|z_2|^4 + |z_3|^4)z_1 - f_1(|z_2|^2 + |z_3|^2)|z_1|^2 z_1 \\ &\quad - f_2|z_2|^2|z_3|^2 z_1 - g\bar{z}_1\bar{z}_2^2\bar{z}_3^2, \qquad (5.71)\\ \frac{dz_2}{dt} &= \mu z_2 - b|z_2|^2 z_2 - c(|z_3|^2 + |z_1|^2)z_2 - d_1|z_2|^4 z_2 \\ &\quad -d_2(|z_3|^4 + |z_1|^4)z_2 - f_1(|z_3|^2 + |z_1|^2)|z_2|^2 z_2 \\ &\quad - f_2|z_3|^2|z_1|^2 z_2 - g\bar{z}_2\bar{z}_3^2\bar{z}_1^2, \qquad (5.72)\end{aligned}$$

$$\frac{dz_3}{dt} = \mu z_3 - b|z_3|^2 z_3 - c(|z_1|^2 + |z_2|^2)z_3 - d_1|z_3|^4 z_3$$
$$-d_2(|z_1|^4 + |z_2|^4)z_3 - f_1(|z_1|^2 + |z_2|^2)|z_3|^2 z_3$$
$$-f_2|z_1|^2|z_2|^2 z_3 - g\bar{z}_3\bar{z}_1^2\bar{z}_2^2, \tag{5.73}$$

where d_1, d_2, f_1, f_2 and g are real constants The solutions to these equations are:

(i) the trivial solution $z_1 = z_2 = z_3 = 0$;
(ii) rolls, e.g $Re(z_1) = R_0$, $Im(z_1) = 0$, $z_2 = z_3 = 0$, with

$$0 = \mu - bR_0^2 - d_1 R_0^4; \tag{5 74}$$

(iii) patchwork quilt, for example, $Re(z_1) = Re(z_2) = R_0$, $Im(z_1) = Im(z_2) = 0$, $z_3 = 0$, with

$$0 = \mu - (b + c)R_0^2 - (d_1 + d_2 + f_1)R_0^4 = 0; \tag{5.75}$$

(iv) hexagons, $z_1 = R_0 e^{i\phi_1}$, $z_2 = R_0 e^{i\phi_2}$, $z_3 = R_0 e^{i\phi_3}$, with

$$\mu - (b + 2c)R_0^2 - (d_1 + 2d_2 + 2f_1 + f_2 + g)R_0^4 = 0, \tag{5 76}$$
$$\phi_1 + \phi_2 + \phi_3 = 0; \tag{5 77}$$

(v) regular triangles, $z_1 = R_0 e^{i\phi_1}$, $z_2 = R_0 e^{i\phi_2}$, $z_3 = R_0 e^{i\phi_3}$, with

$$\mu - (b + 2c)R_0^2 - (d_1 + 2d_2 + 2f_1 + f_2 - g)R_0^4 = 0, \tag{5 78}$$
$$\phi_1 + \phi_2 + \phi_3 = \frac{\pi}{2} \tag{5 79}$$

The stability of these solutions can be determined from the amplitude equations (5 71)–(5.73) Any one of rolls, hexagons and regular triangles can be stable close to the bifurcation point, though no more than one solution can be stable at onset simultaneously. Full details of the stability calculations can be found in Golubitsky, Swift and Knobloch (1984).

5.5 Roll/stripe solutions

Roll (stripe) solutions fit onto either of the square or hexagonal lattices. By setting $z_1 \equiv z$ and all other amplitudes to zero, we can see that in either case, the amplitude equation for rolls takes the general form

$$\frac{dz}{dt} = \mu z - a|z|^2 z, \tag{5.80}$$

where a is a real constant However, the stability properties of the rolls depend on the lattice, because this determines which other patterns they are in competition with. On the other hand we could choose to investigate a system in only one space

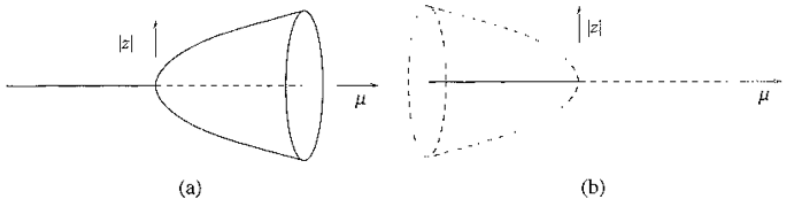


Fig. 5.9. Bifurcation diagrams for the (a) supercritical and (b) subcritical pitchforks of revolution

dimension, where only rolls can arise. The lattice would consist of an array of equally spaced points along a line as shown in Figure 5.2a, and to leading order we could write the solution

$$u(x,t) = z(t)e^{ix} + c.c., \tag{5.81}$$

where now $x \in \mathbb{R}$ and $z \in \mathbb{C}$. The relevant Euclidean symmetries would be translations and reflections in x. Restricting these to the lattice we end up with the reflection $x \to -x$ and the circle of translations modulo 2π, which together generate the group $O(2)$, the group of rotations and reflections of the circle, where the translations are playing the role of the rotations. Applying the equivariance condition would lead directly to equation (5.80). This looks something like the normal form for a pitchfork bifurcation. In fact, since z is complex, we call it a **pitchfork of revolution**, the idea being that the equation holds for z with any phase between 0 and 2π, so we have a whole circle of pitchforks. The bifurcation diagrams are shown in Figure 5.9.

The only solutions to equation (5.80) are the trivial solution $z = 0$ and the roll solution $|z|^2 = \mu/a$. A simple calculation shows that rolls are stable in the supercritical case $a > 0$, and unstable in the subcritical case $a < 0$. The trivial solution is stable for $\mu < 0$ and unstable for $\mu > 0$.

It is easy to adapt the methods we have just been looking at to deal with more unusual situations, such as a lack of reflection symmetry, or other type of anisotropy (where not all directions are equivalent): the symmetry group will no longer be $E(2)$, so you simply apply equivariance under the new symmetry group and see what you get. In the next section we shall consider a situation where there is no reflection symmetry, and there are further examples in the exercises at the end of the chapter.

5.6 The Küppers–Lortz instability

Consider the competition between two sets of rolls in a rotating system. We will write the solution in the form

$$u(x,t) = z_1(t)e^{ix} + z_2(t)e^{i\tilde{x}} + c.c. + h.o.t., \tag{5.82}$$

where $x, \tilde{x} \in \mathbb{R}$ are such that the $\tilde{x}$ direction is rotated anticlockwise through an angle, θ, relative to the x direction and the two variables have the same scaling. The solutions lie on a rhombic dual lattice defined by the unit vectors in the x and $\tilde{x}$ directions. Using translation symmetry in two orthogonal directions we can deduce that the amplitude equations for z_1 and z_2 must take the form

$$\frac{dz_1}{dt} = \mu z_1 - \gamma |z_1|^2 z_1 - \alpha |z_2|^2 z_1, \tag{5.83}$$

$$\frac{dz_2}{dt} = \nu z_2 - \delta |z_2|^2 z_2 - \beta |z_1|^2 z_2. \tag{5.84}$$

Let us assume that the system is isotropic so that the linear growth rate of the two modes is the same, $\mu = \nu$, and the coefficients of the self-interaction terms $|z_1|^2 z_1$ and $|z_2|^2 z_2$ are the same $\gamma = \delta$. We will be interested in the case where rolls are a stable solution of the equations, so the coefficient of the self-interaction term must be negative, and we can set $\gamma = \delta = 1$ by rescaling z_1 and z_2. Since the system is rotating, however, there is no reflection symmetry in the system, so the cross-interaction terms $|z_2|^2 z_1$ and $|z_1|^2 z_2$ need not have the same coefficient. The amplitude equations thus reduce to

$$\frac{dz_1}{dt} = \mu z_1 - |z_1|^2 z_1 - \alpha |z_2|^2 z_1, \tag{5.85}$$

$$\frac{dz_2}{dt} = \mu z_2 - |z_2|^2 z_2 - \beta |z_1|^2 z_2. \tag{5.86}$$

The solutions to these equations are the trivial solution, $z_1 = z_2 = 0$, x-rolls ($|z_1|^2 = \mu$, $z_2 = 0$), $\tilde{x}$-rolls ($z_1 = 0$, $|z_2|^2 = \mu$) and rectangles ($|z_1|^2/(1-\alpha) = |z_2|^2/(1-\beta) = \mu/(1-\alpha\beta)$). Consider the case $\mu > 0$, $(1-\alpha)(1-\beta) < 0$, where rectangles cannot occur. The trivial solution is unstable since $\mu > 0$. The stability of the x-rolls can be investigated by setting $z_1 = \sqrt{\mu}(1+a)$, $z_2 = b$ where $|a| \ll 1$ and $|b| \ll \sqrt{\mu}$. Linearising in the perturbations a and b we have

$$\frac{d(a+\bar{a})}{dt} = -2\mu(a+\bar{a}), \tag{5.87}$$

$$\frac{d(a-\bar{a})}{dt} = 0, \tag{5.88}$$

$$\frac{db}{dt} = \mu(1-\beta)b \tag{5.89}$$

So x-rolls will be stable if $\beta > 1$ (since we require $\mu > 0$ already for their existence). Similarly $\tilde{x}$-rolls will be stable if $\alpha > 1$.

If we pick $\alpha > 1$ and $\beta < 1$, then x-rolls will lose stability to $\tilde{x}$-rolls. So the original rolls are unstable to new rolls rotated through some angle, θ, anticlockwise from the originals. Once established, these rolls can themselves become unstable to yet more rolls, rotated through a further angle, θ, and so on. Rolls grow, then

disappear and are replaced by rolls rotated anticlockwise through the angle, θ, indefinitely. The third set of rolls do not fit on the original rhombic lattice, but if the original rolls have entirely died away by the time the third set start to grow, then perhaps it is not too bad an approximation to consider a new rhombic lattice on which only the second and third sets fit.

If $\alpha < 1$ and $\beta > 1$, then $\tilde{x}$-rolls will lose stability to x-rolls, and the rotating sets of rolls will proceed in the clockwise direction.

The possibility of this cyclic loss of stability of rolls to rotated rolls was noted by Küppers and Lortz (1969) in their stability calculations for rolls in rotating convection. The angle between each successive set of rolls was about 58°. The rolls turn about the vertical axis in the same direction as the applied rotation. The Busse–Heikes heteroclinic cycle discussed in Section 4.5 was originally devised as a model for the Küppers–Lortz instability: each fixed point in the cycle represents a set of rolls at an angle of $2\pi/3$ to the previous one.

Figure 5.10 shows a numerical simulation of a Küppers–Lortz instability in rotating convection modelled by the equation

$$\frac{\partial w}{\partial t} = \mu w - (1+\nabla^2)^2 w - w(w^2 + |\nabla w|^2) + \frac{2}{5}\nabla w \quad \nabla(w^2 + |\nabla w|^2) + V\hat{z}\cdot\nabla\times(|\nabla w|^2\nabla w), \tag{5.90}$$

where $w(\boldsymbol{x}, t)$ is the vertical velocity of the convecting fluid, $\hat{z}$ is a vertical unit vector, V measures the strength of the rotation about a vertical axis, and μ is the reduced Rayleigh number measuring the distance from onset of the convection (introduced in Chapter 1). This is a simplification of a model derived by Ponty, Passot and Sulem (1997), and has been studied by Pollicott, Matthews and Cox (2003).

We consider a solution to equation (5.90) of the form

$$w(\boldsymbol{x}, t) = z_1(t)e^{i\boldsymbol{x}} + z_2(t)e^{i\tilde{\boldsymbol{x}}} + c.c. + h.o.t., \tag{5.91}$$

where $\boldsymbol{x} = (\cos\theta_0, \sin\theta_0)\ \boldsymbol{x}$ and $\tilde{\boldsymbol{x}} = (\cos(\theta_0+\theta), \sin(\theta_0+\theta))\ \boldsymbol{x}$, so that the $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ directions lie at angles θ_0 and $\theta_0+\theta$ to the x_1 axis respectively, where $\boldsymbol{x} = (x_1, x_2)$. If this is substituted into equation (5.90), the amplitude equations for z_1 and z_2 turn out to be

$$\frac{\mathrm{d}z_1}{\mathrm{d}t} = \mu z_1 - 4|z_1|^2 z_1 - 2\left(4 + V\sin 2\theta - \frac{2}{5}(1-\cos 2\theta)\right)|z_2|^2 z_1, \tag{5.92}$$

$$\frac{\mathrm{d}z_2}{\mathrm{d}t} = \mu z_2 - 4|z_2|^2 z_2 - 2\left(4 - V\sin 2\theta - \frac{2}{5}(1-\cos 2\theta)\right)|z_1|^2 z_2, \tag{5.93}$$

to leading order close to the onset of convection. The method for deriving amplitude equations such as these from governing partial differential equations will be discussed in Chapter 7; for now we will simply accept that it can be done. After rescaling both z_1 and z_2 by a factor of two these are just equations (5.85) and (5.86) with

$$\alpha = \frac{1}{2}\left(4 + V\sin 2\theta - \frac{2}{5}(1-\cos 2\theta)\right), \tag{5.94}$$

$$\beta = \frac{1}{2}\left(4 - V\sin 2\theta - \frac{2}{5}(1-\cos 2\theta)\right). \tag{5.95}$$

Considering perturbations to the z_1-rolls, the growth rate of z_2 perturbations is

$$\sigma = \mu\left(-1 + \frac{1}{2}V\sin 2\theta + \frac{1}{5}(1-\cos 2\theta)\right), \tag{5.96}$$

from equation (5.89). This is maximised when $\tan 2\theta = -5V/2$, and the condition for instability, $\sigma > 0$, then becomes

$$\mu(-8 + \sqrt{4 + 25V^2}) > 0 \tag{5.97}$$

This is achieved for $\mu > 0$ if $V^2 > 60/25$: in other words if the rotation speed is high enough. The progress of the instability is shown in Figure 5.10.

5.7 Hopf bifurcation on a one-dimensional lattice

In this section we will introduce oscillatory behaviour, looking at the simplest oscillating lattice patterns: one-dimensional waves.

Imagine we want to describe a one-dimensional oscillatory pattern that sets in at the bifurcation point with clearly defined spatial and temporal periods. One way of writing such a pattern close to onset is

$$u(x,t) = z_1(t)e^{i(x-t)} + z_2(t)e^{-i(x+t)} + c.c. + h.o.t., \tag{5.98}$$

where $z_1, z_2 \in \mathbb{C}$, and where the spatial and temporal periods at onset have been set to 2π, by scaling x and t if necessary.

We are using the same one-dimensional spatial lattice as we did for the periodic stripe solution, but now we have added in some oscillatory behaviour. We need both the $e^{i(x-t)}$ and $e^{-i(x-t)}$ terms in our general solution in order to account for all the possible combinations of spatial and temporal behaviour.

When z_1 is nonzero and $z_2 = 0$, the solution $u(x,t)$ represents a plane wave travelling to the right (towards positive x), whereas when z_2 is nonzero and $z_1 = 0$ we have a plane wave travelling to the left. If $z_1 = \pm z_2$ we have a standing wave. Examples of standing and travelling waves are shown in Figure 5.11.

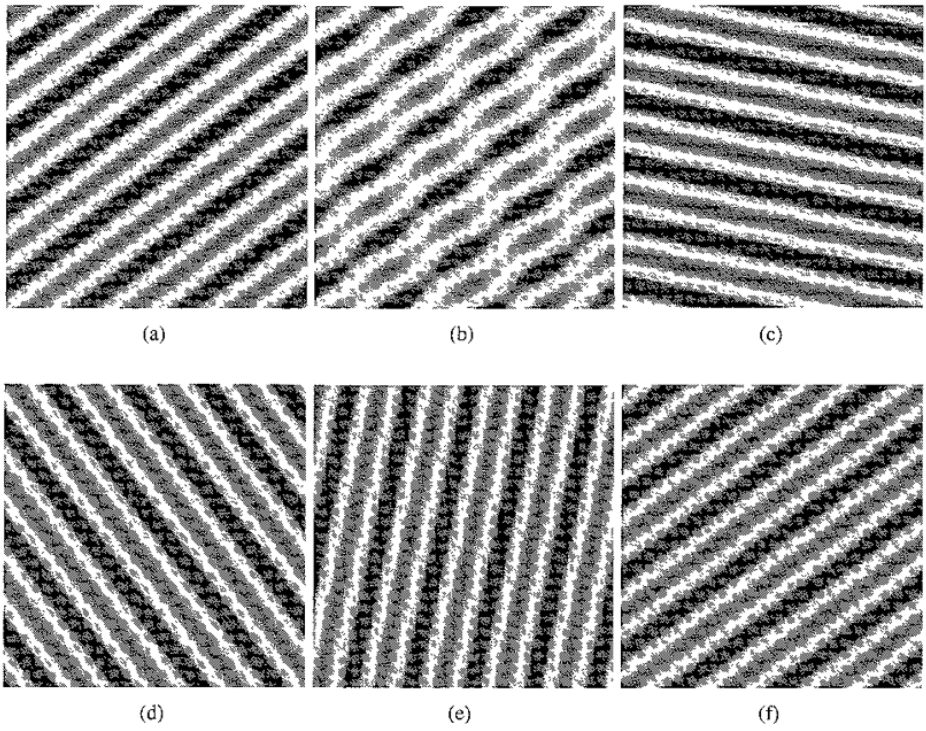


Fig 5.10 Simulation of a Küppers–Lortz instability in rotating convection modelled by equation (5 90) with $\mu = 0\,2$ and $V = 2$ The domain has sides of length 10π. (a) At $t = 0$, the roll wavevector is at $\theta = -53°$ to the x_1-axis (b) The rolls are unstable, and perturbations grow (pictured at $t = 672$) (c) By $t = 1550$, the original rolls have lost stability to new rolls with wavevector at $\theta = -101\ 31°$ to the x_1-axis The process repeats, and rolls emerge at (d) $\theta = -143\ 12°$ ($t = 3875$) and then (e) $\theta = -191\ 31°$ ($t = 10075$) (f) At $t = 15\,000$ the original roll pattern reappears Since the rolls must fit exactly into the finite integration domain, their wavelength varies slightly through the cycle, as does the angle of rotation between successive patterns Pictures courtesy of and ©Sarah Pollicott, University of Nottingham, 2003.

Now let us assume that the system we are working in has symmetry under the reflection

$$m : x \to -x,\ (z_1, z_2) \to (z_2, z_1) \tag{5 99}$$

and translations in x

$$p : x \to x + p,\ (z_1, z_2) \to (z_1 e^{-ip}, z_2 e^{ip}). \tag{5.100}$$

As we saw in the case of the steady bifurcation on a one-dimensional lattice, this system has $O(2)$ symmetry. We have already looked at the Hopf bifurcation with

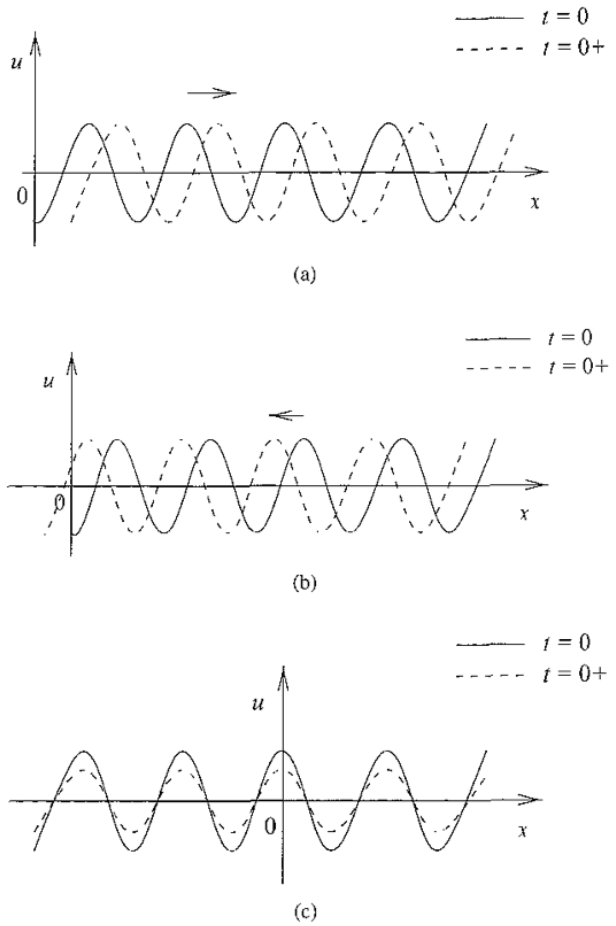


Fig. 5.11. Examples of travelling and standing waves, as described by equation (5.98): (a) z_1 constant, $z_2 = 0$, (b) $z_1 = 0$, z_2 constant and (c) $z_1 = z_2$ =constant. In each case the solid line shows the solution at time $t = 0$ and the dotted line shows the solution at a slightly later time $t = 0+$

$O(2)$ symmetry in Section 4.4.1. The representation is the same here: we have two copies of $\mathbb{C}$, one corresponding to z_1 and the other to z_2 – it is the time dependence (the oscillatory nature of the bifurcation) that makes them 'different'. Comparing the symmetries we have here with those given in equations (4.124)–(4.126) we see that the translation p takes the rôle of the rotation in equation (4.124), and that so far we are missing a phase-shift symmetry. In fact our system should be unchanged under a shift of the origin in time if we assume the usual case where there are no special instants of time – in other words, all things being equal, the hypothetical experiment you are performing doesn't particularly care whether it

is Tuesday lunchtime or Saturday night. So there should be symmetry under $t \to t + \theta$ or $(z_1, z_2) \to (z_1 e^{i\theta}, z_2 e^{i\theta})$, which is just the phase-shift symmetry from equation (4.126).

The travelling waves we have described are exactly the rotating waves from Section 4.4.1. The right-travelling wave, $z_1 e^{i(x-t)} + c.c.$, has isotropy subgroup $\widetilde{SO(2)}$, generated by $(\theta, \theta) : (x, t) \to (x + \theta, t + \theta)$, and the left-travelling wave, $z_2 e^{-i(x+t)} + c.c.$, has the conjugate isotropy subgroup generated by $(-\theta, \theta)$. The standing waves, $z_1(e^{i(x-t)} + e^{-i(x+t)}) + c.c.$, have isotropy subgroup $\mathbb{Z}_2 \times \mathbb{Z}_2^c$ just as in Section 4.4.1.

We will now derive the amplitude equations for z_1 and z_2 to see how the travelling and standing waves evolve over time. Equivariance under the translations in x and t lead to the equation

$$\frac{dz_1}{dt} = (\mu + i\nu)z_1 - (\alpha + i\beta)|z_1|^2 z_1 - (\gamma + i\delta)|z_2|^2 z_1, \tag{5.101}$$

up to cubic order, where μ, ν, α, β, γ and δ are real constants. We can now deduce the equation for z_2 using the reflection symmetry $m : (z_1, z_2) \to (z_2, z_1)$ to find

$$\frac{dz_2}{dt} = (\mu + i\nu)z_2 - (\alpha + i\beta)|z_2|^2 z_2 - (\gamma + i\delta)|z_1|^2 z_2 \tag{5.102}$$

The parameter ν is the **frequency detuning**, and will in general depend on μ. The requirement that the solutions have period 2π at the bifurcation point $\mu = 0$ means that we must set $\nu = 0$ there. Away from the bifurcation point ν will typically be nonzero, but it can be removed from the amplitude equations using the transformation $z_j \to z_j e^{i\nu t}$ for $j = 1, 2$, which is equivalent to changing the carrier wave frequency to $-1 + \nu$. To cubic order, then, the amplitude equations become

$$\frac{dz_1}{dt} = \mu z_1 - (\alpha + i\beta)|z_1|^2 z_1 - (\gamma + i\delta)|z_2|^2 z_1, \tag{5.103}$$

$$\frac{dz_2}{dt} = \mu z_2 - (\alpha + i\beta)|z_2|^2 z_2 - (\gamma + i\delta)|z_1|^2 z_2. \tag{5.104}$$

We can put these equations into amplitude and phase form by writing $z_1 = Re^{i\phi}$ and $z_2 = Se^{i\psi}$, where $R, S > 0$ and $0 \le \phi, \psi < 2\pi$, and separating out the real and imaginary parts of the equations to get

$$\frac{dR}{dt} = \mu R - \alpha R^3 - \gamma S^2 R, \tag{5.105}$$

$$\frac{d\phi}{dt} = -\beta R^2 - \delta S^2, \tag{5.106}$$

$$\frac{dS}{dt} = \mu S - \alpha S^3 - \gamma R^2 S, \tag{5.107}$$

$$\frac{d\psi}{dt} = -\beta S^2 - \delta R^2. \tag{5.108}$$

Looking for solutions of constant amplitude, so that $dR/dt = dS/dt = 0$, we find

(i) the trivial solution $z_1 = z_2 = 0$ ($R = S = 0$);
(ii) right-travelling waves

$$R^2 = \frac{\mu}{\alpha}, \quad \frac{d\phi}{dt} = -\frac{\beta\mu}{\alpha}, \quad S = 0; \tag{5.109}$$

(iii) left-travelling waves

$$S^2 = \frac{\mu}{\alpha}, \quad \frac{d\psi}{dt} = -\frac{\beta\mu}{\alpha}, \quad R = 0; \tag{5.110}$$

(iv) standing waves

$$R^2 = S^2 = \frac{\mu}{(\alpha+\gamma)}, \quad \frac{d\phi}{dt} = \frac{d\psi}{dt} = -\frac{\mu(\beta+\delta)}{(\alpha+\gamma)} \tag{5.111}$$

Both travelling and standing waves bifurcate from the trivial solution at $\mu = 0$ as predicted by the equivariant Hopf theorem. Note that there are no constant amplitude solutions with unequal, but nonzero, values of R^2 and S^2 (referred to as 'general points' in Table 4.4) that appear as solutions of the truncated amplitude equations

We can investigate the stability of the travelling wave and standing wave solutions using the truncated amplitude equations (5.103) and (5.104). First note that the equations for $d\phi/dt$ and $d\psi/dt$ depend only on R and S, so that if R and S are stable to perturbations, then so will ϕ and ψ be. That means we only need to look at the stability of the amplitudes R and S to perturbations.

Clearly the trivial solution $R = S = 0$ is stable for $\mu < 0$ and unstable for $\mu > 0$.

We perturb the right-travelling wave $R^2 = \mu/\alpha > 0$, $S = 0$ by setting $R = \sqrt{\mu/\alpha}(1 + r)$, $S = s$, where $|r|, |s| \ll 1$. Linearising in the perturbations gives

$$\frac{dr}{dt} = -2\mu r, \tag{5.112}$$

$$\frac{ds}{dt} = \mu\left(1 - \frac{\gamma}{\alpha}\right)s. \tag{5.113}$$

So we have stability if $\mu > 0$ (which implies $\alpha > 0$) and $\gamma > \alpha$. Left-travelling

waves have the same stability conditions, as can be seen by interchanging R and S. In fact left- and right-travelling waves must have the same existence and stability properties, because they lie on the same group orbit and have conjugate isotropy subgroups

The standing wave solution is given by $R^2 = S^2 = \mu/(\alpha + \gamma) > 0$. We perturb it by setting $R = \sqrt{\mu/(\alpha+\gamma)}(1+r)$, $S = \sqrt{\mu/(\alpha+\gamma)}(1+s)$, with $|r|, |s| \ll 1$. Now linearising in r and s gives

$$\frac{dr}{dt} = -\frac{2\alpha\mu r}{(\alpha+\gamma)} - \frac{2\gamma\mu s}{(\alpha+\gamma)}, \tag{5.114}$$

$$\frac{ds}{dt} = -\frac{2\alpha\mu s}{(\alpha+\gamma)} - \frac{2\gamma\mu r}{(\alpha+\gamma)} \tag{5.115}$$

Adding and subtracting the equations gives

$$\frac{d}{dt}(r+s) = -2\mu(r+s), \tag{5.116}$$

$$\frac{d}{dt}(r-s) = -\frac{2\mu(\alpha-\gamma)}{(\alpha+\gamma)}(r-s). \tag{5.117}$$

So we have stability for $\mu > 0$ and $\alpha > \gamma$

You have probably noticed that apart from a few changes of notation this is exactly the same calculation as we did for the bifurcations in a square box and on a square lattice, where now travelling waves are the equivalent of rolls, and standing waves are the equivalent of squares, and we can identify α and γ with a_1 and a_2 of equations (5.40) and (5.41) respectively. The various possible bifurcation diagrams for this system can now be read from Figure 4.10

Exercises

5.1 Show that the action of $D_4 \ltimes T^2$ given by the matrices in lines (5.18) and (5.19) is absolutely irreducible.
Extension Show that the fundamental representation of $D_6 \ltimes T^2$ is also absolutely irreducible, using a similar method

5.2 Determine the criteria for stability of the patchwork quilt solution of the quintic order amplitude equations (5.71)–(5.73) for the steady bifurcation on a hexagonal lattice in the presence of midplane reflection symmetry
Extension Do the same for the regular triangle, roll and hexagon solutions

5.3 Work out the amplitude equations for the steady bifurcation on a square lattice when the system is weakly anisotropic so that the x_1 and x_2 directions are not equivalent What happens to the square solutions?

5.4 What happens to the steady bifurcation on a hexagonal lattice if the reflection m_v is weakly broken? Are rectangles still a solution? What about hexagons?

5.5 Work out the amplitude equations for a Hopf bifurcation on a square lattice, where the pattern takes the form

$$u(\boldsymbol{x}, t) = z_1 e^{i(x_1 - t)} + z_2 e^{-i(x_1 + t)} + z_3 e^{i(x_2 - t)} + z_4 e^{-i(x_2 + t)} + c.c.,$$

where $z_j \in \mathbb{C}$ for $j = 1, 2, 3, 4$, to leading order

6

Superlattices, hidden symmetries and other complications

There are many more symmetric spatially extended patterns than we have seen so far, for example superlattice patterns related to new representations of the lattice, patterns that involve modes of different spatial periods or that change their spatial period at bifurcation, and quasipatterns that look highly symmetrical, but don't fit on any lattice. There are also 'hidden' symmetries lurking in seemingly straightforward bifurcation problems.

6.1 Superlattice patterns

Up to this point we have always used the fundamental representation of the symmetry group for the lattice under consideration. However this is not the only choice. Consider the square dual lattice generated by the wavevectors $\boldsymbol{k}_1 = (1, 0)$ and $\boldsymbol{k}_2 = (0, 1)$ shown in Figure 6.1. If the critical wavenumber is $k_c = 1$ then we have the fundamental representation of $D_4 \ltimes I^2$ that we used before. On the other hand if $k_c = \sqrt{5}$, the critical circle intersects eight lattice points and we have another absolutely irreducible representation

$$\begin{aligned} u(x, y, t) = {} & z_1(t)e^{i(2x+y)} + z_2(t)e^{i(-x+2y)} + z_3(t)e^{i(x+2y)} \\ & + z_4(t)e^{i(-2x+y)} + c.c \end{aligned} \tag{6.1}$$

to leading order, where $(x, y) \in \mathbb{R}^2$ are Cartesian coordinates and the amplitudes z_j are complex, so the representation acts on $\mathbb{C}^4$. There are also higher-dimensional reducible representations such as the one for $k_c = 5$ that intersects twelve lattice points.

In fact Crawford (1994) and Dionne, Silber and Skeldon (1997) point out that there is a countable infinity of absolutely irreducible representations of $D_4 \ltimes I^2$ that take the form

$$\begin{aligned} u(x, y, t) = {} & z_1(t)e^{i(\alpha x+\beta y)} + z_2(t)e^{i(-\beta x+\alpha y)} + z_3(t)e^{i(\beta x+\alpha y)} \\ & + z_4(t)e^{i(-\alpha x+\beta y)} + c.c\ , \end{aligned} \tag{6.2}$$

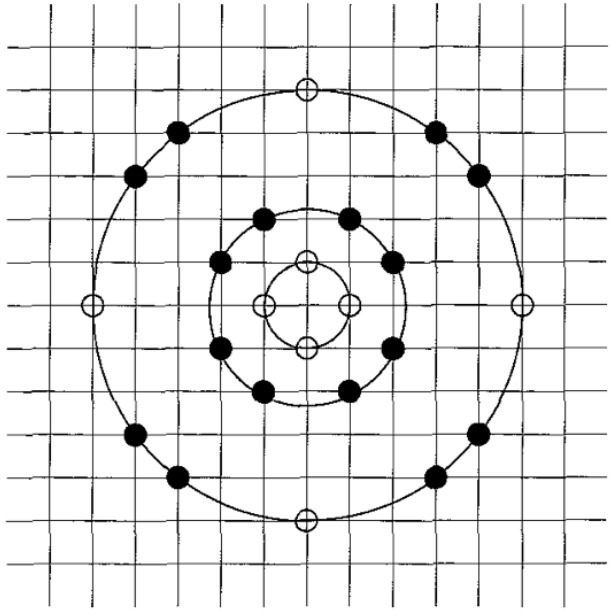


Fig 6.1. Intersections of the critical circle with the square dual lattice $\mathcal{L}^*$ generated by $\boldsymbol{k}_1 = (1, 0)$ and $\boldsymbol{k}_2 = (0, 1)$ for $k_c = 1$, $k_c = \sqrt{5}$ and $k_c = 5$. Filled and empty dots on the critical circle $k_c = 5$ belong to different irreps making up the reducible representation. After Dionne, Silber and Skeldon (1997)

where α and β are positive integers. These representations are **translation-free** if α and β are relatively prime and not both odd: that is, there are no nontrivial translations that act trivially on equation (6.2), so the neutral modes cannot be supported on a finer lattice. It is helpful to note for later that exactly one of α and β must be odd.

For a steady bifurcation, we can work out the action of $D_4 \ltimes T^2$ on $\mathbb{C}^4$ given by equation (6.2) by considering the generators:

(i) rotation through $\pi/2$: $\rho : (z_1, z_2, z_3, z_4) \to (\bar{z}_2, z_1, \bar{z}_4, z_3)$;
(ii) reflection: $m_y : (z_1, z_2, z_3, z_4) \to (\bar{z}_4, \bar{z}_3, \bar{z}_2, \bar{z}_1)$;
(iii) translations: $\boldsymbol{p} : (z_1, z_2, z_3, z_4) \to (z_1 e^{-i(\alpha p_1 + \beta p_2)}, z_2 e^{i(\beta p_1 - \alpha p_2)}, z_3 e^{-i(\beta p_1 + \alpha p_2)}, z_4 e^{i(\alpha p_1 - \beta p_2)})$

The isotropy subgroups with one-dimensional fixed-point subspace are given in Table 6.1, and the corresponding solutions are illustrated in Figure 6.2. There are new branches in addition to the simple square and roll branches found under the fundamental representation, and furthermore the squares and rolls do not appear to have the same isotropy subgroups that they had previously – some of the expected symmetries are missing, which is rather surprising. So what is going on here? It

Table 6.1 *The axial isotropy subgroups of $\mathcal{D}_4 \ltimes T^2$ under the representation (6.2), their generators and fixed-point subspaces. Adapted from Dionne, Silber and Skeldon (1997)*

Branch	Σ	Generators	Fix(Σ)
Super squares	D_4	ρ, m_y	$z_1 = z_2 = z_3 = z_4 \in \mathbb{R}$
Anti-squares	$\widetilde{D}_4$	$\rho, (m_y, (\pi, \pi))$	$z_1 = z_2 = -z_3 = -z_4 \in \mathbb{R}$
Rolls	$\mathbb{Z}_2^c \ltimes S_1^1$	$\rho^2, \left(e, \left(\frac{2\pi\alpha}{\alpha^2+\beta^2}, \frac{2\pi\beta}{\alpha^2+\beta^2}\right)\right)$, $(e, (\beta s, -\alpha s))$, $s \in \mathbb{R}$	$z_1 \in \mathbb{R}, z_2 = z_3 = z_4 = 0$
Simple squares	$\mathbb{Z}_4 \ltimes S_{1,2}$	$\rho, \left(e, \left(\frac{2\pi\alpha}{\alpha^2+\beta^2}, \frac{2\pi\beta}{\alpha^2+\beta^2}\right)\right)$, $\left(e, \left(\frac{-2\pi\beta}{\alpha^2+\beta^2}, \frac{2\pi\alpha}{\alpha^2+\beta^2}\right)\right)$	$z_1 = z_2 \in \mathbb{R}, z_3 = z_4 = 0$
Rhombs 1	$\mathcal{D}_2 \ltimes S_{1,3}$	$\rho^2, m_y \rho^3$, $\left(e, \left(\frac{2\pi\alpha}{\alpha^2-\beta^2}, \frac{-2\pi\beta}{\alpha^2-\beta^2}\right)\right)$, $\left(e, \left(\frac{-2\pi\beta}{\alpha^2-\beta^2}, \frac{2\pi\alpha}{\alpha^2-\beta^2}\right)\right)$	$z_1 = z_3 \in \mathbb{R}, z_2 = z_4 = 0$
Rhombs 2	$D_2 \ltimes S_{1,4}$	$\rho^2, m_y, \left(e, \left(\frac{\pi}{\alpha}, \frac{\pi}{\beta}\right)\right)$, $\left(e, \left(\frac{-\pi}{\alpha}, \frac{\pi}{\beta}\right)\right)$	$z_1 = z_4 \in \mathbb{R}, z_2 = z_3 = 0$

turns out that the 'missing' symmetries are just hidden. In this context a hidden symmetry is one that is not present in the holohedry of the lattice, so does not preserve the lattice, but which leaves invariant some fixed-point subspace of the bifurcation problem restricted to the lattice. Hidden symmetries will be discussed in more detail in Section 6.6. Dionne and Golubitsky (1992) showed that patterns that have isotropy subgroups containing nontrivial translations actually fit onto a finer lattice (coarser dual lattice), and within the context of the finer lattice the isotropy subgroup contains all the symmetries of the patterns, so there are no hidden symmetries. It can be seen in Figure 6.2 that rolls, simple squares and rhombs are all periodic on the short scale $2\pi/k_c = 2\pi(\alpha^2 + \beta^2)^{-\frac{1}{2}}$, while super squares and anti-squares have the longer spatial period 2π in both x and y directions, since the basic wavevectors of the dual lattice have unit length. However, both super square and anti-square planforms show features on the short scale $1/k_c$ within each basic period: patterns like these have come to be known as **superlattice patterns**.

Rolls with $z_1 \in \mathbb{R}$, $z_2 = z_3 = z_4 = 0$ fit onto the finer one-dimensional lattice defined by the wavevector $\boldsymbol{k} = (\alpha, \beta)$. The nontranslational part of the isotropy subgroup on the original lattice was $\mathbb{Z}_2^c$, generated by ρ^2. On the finer lattice the isotropy subgroup is $\mathcal{D}_2 \ltimes S^1$, where $\mathcal{D}_2$ is generated by ρ^2 and a reflection with

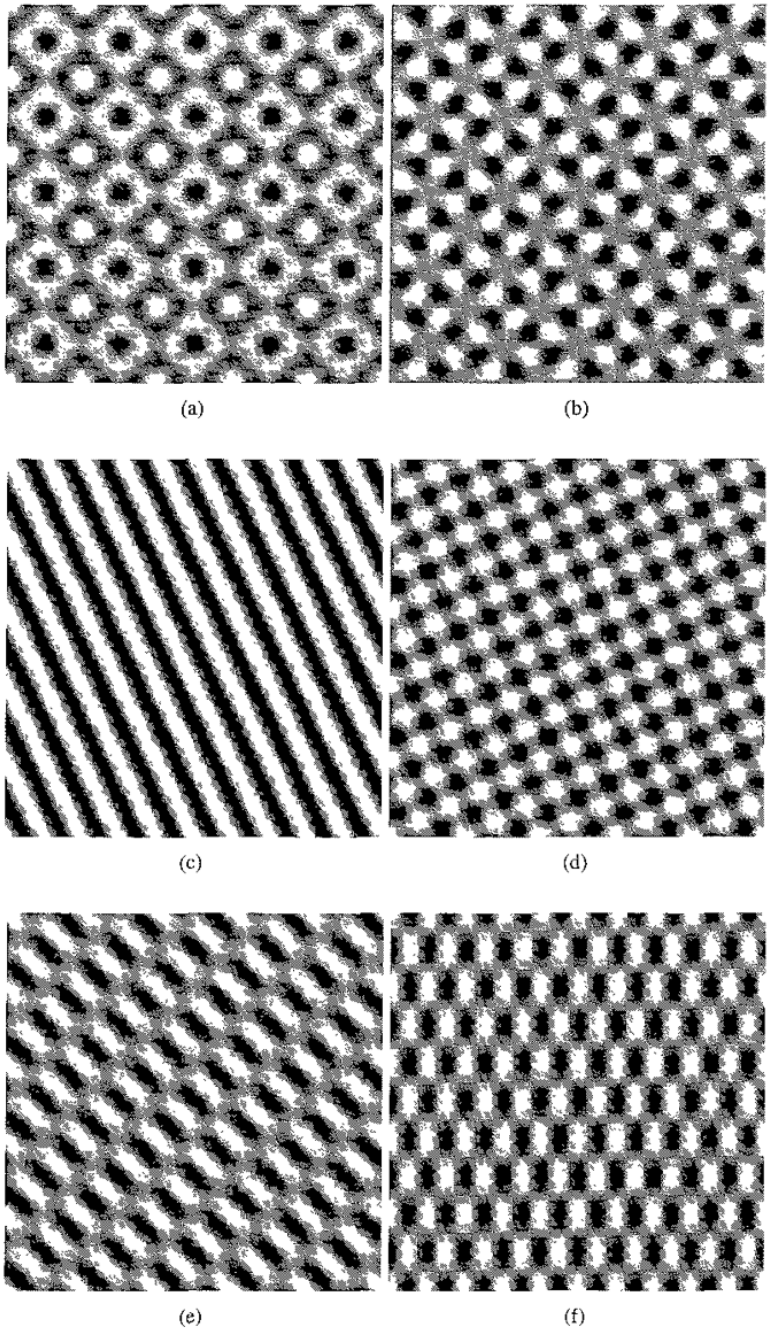


Fig. 6.2 Solutions of the form $u(x, y) = z_1 e^{i(2x+y)} + z_2 e^{i(-x+2y)} + z_3 e^{i(x+2y)} + z_4 e^{i(-2x+y)} + c.c.$ shown in greyscale on the square domain $x, y \in [0, 10\pi)$ with amplitudes as given in Table 6.1 (all nonzero z_j are set to 1): (a) super squares, (b) anti-squares, (c) rolls, (d) simple squares, (e) rhombs 1 and (f) rhombs 2

axis $(-\beta, \alpha)$ given by

$$\begin{pmatrix} x \\ y \end{pmatrix} \to \frac{1}{\alpha^2+\beta^2} \begin{pmatrix} \beta^2-\alpha^2 & -2\alpha\beta \\ -2\alpha\beta & \alpha^2-\beta^2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \tag{6.3}$$

$$(z_1, z_2, 0, 0) \to (\bar{z}_1, z_2, 0, 0), \tag{6.4}$$

that was previously hidden as it is not a symmetry of the original lattice. Similarly, simple squares with $z_1 = z_2 \in \mathbb{R}$, $z_3 = z_4 = 0$ fit onto the finer lattice defined by the wavevectors $\boldsymbol{k}_1 = (\alpha, \beta)$ and $\boldsymbol{k}_2 = (-\beta, \alpha)$, where their isotropy subgroup is $\mathcal{D}_4$, generated by ρ and the reflection with axis $(-\beta, \alpha)$ described above. Before, the nontranslational part of the isotropy subgroup was just $\mathbb{Z}_4$, as the reflection was hidden. The whole subspace, $(z_1, z_2, 0, 0) : z_1, z_2 \in \mathbb{C}$, of the bifurcation problem is invariant under this hidden reflection, and z_1 and z_2 rolls also fit on the finer square lattice. Rhombs 1 ($z_1 = z_3 \in \mathbb{R}$, $z_2 = z_4 = 0$) fit onto the finer rhombic lattice defined by wavevectors $\boldsymbol{k}_1 = (\alpha, \beta)$ and $\boldsymbol{k}_2 = (\beta, \alpha)$, and rhombs 2 ($z_1 = z_4 \in \mathbb{R}$, $z_2 = z_3 = 0$) onto the finer rhombic lattice defined by $\boldsymbol{k}_1 = (\alpha, \beta)$ and $\boldsymbol{k}_2 = (-\alpha, \beta)$. Both have symmetry group $\mathcal{D}_2$ on the finer lattice, and no hidden symmetries are involved. Calculating the amplitude equations and the stability criteria for each of the patterns is rather a long job, so we won't go into the details here, but they can be found in Dionne, Silber and Skeldon (1997). The equation for z_1 turns out to be

$$\frac{\mathrm{d}z_1}{\mathrm{d}t} = \mu z_1 + z_1(a_1|z_1|^2 + a_2|z_2|^2 + a_3|z_3|^2 + a_4|z_4|^2) + O(|z|^5), \tag{6.5}$$

where the a_j are real and μ is the real bifurcation parameter. The equations for the remaining z_j can be found from equation (6.5) using equivariance. It turns out that the hidden symmetries do not restrict the form of the evolution equations in this case, though we will see in Section 6.6 that hidden symmetries can often have that effect. The stability criteria for all of the planforms would take rather a long time to summarise, but there are some general statements that can be made subject to certain nondegeneracy conditions on the a_j. In particular, any one of the axial planforms can be stable, and if all the axial solution branches bifurcate supercritically then at least one of them will be stable.

There is also a countable infinity of twelve-dimensional representations of $\mathcal{D}_6 \ltimes T^2$ of the form

$$u(x, y, t) = \sum_{j=1}^{6} z_j(t) e^{i\boldsymbol{\kappa}_j \cdot \boldsymbol{x}} + \text{c.c.}, \quad z_j \in \mathbb{C} \tag{6.6}$$

where the dual lattice is generated by the wavevectors $\boldsymbol{k}_1 = (0, 1)$ and $\boldsymbol{k}_2 =$

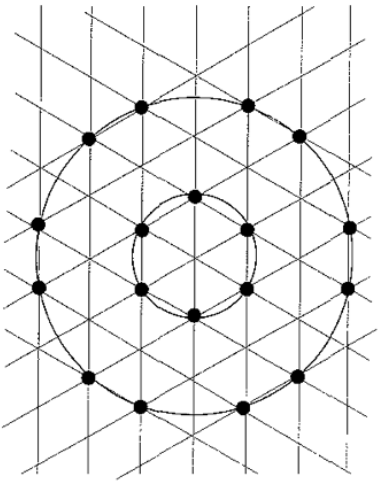


Fig. 6.3 Intersections of the critical circle with the hexagonal dual lattice $\mathcal{L}^*$ generated by $\boldsymbol{k}_1 = (0, 1)$ and $\boldsymbol{k}_2 = (\sqrt{3}/2, -1/2)$, for $k_c = 1$ and $k_c = \sqrt{7}$. After Dionne, Silber and Skeldon (1997)

$(\sqrt{3}/2, -1/2)$, and the critical circle intersects the dual lattice at

$$\boldsymbol{\kappa}_1 = \alpha\boldsymbol{k}_1 + \beta\boldsymbol{k}_2, \tag{6.7}$$

$$\boldsymbol{\kappa}_2 = (-\alpha + \beta)\boldsymbol{k}_1 - \alpha\boldsymbol{k}_2, \tag{6.8}$$

$$\boldsymbol{\kappa}_3 = -\beta\boldsymbol{k}_1 + (\alpha - \beta)\boldsymbol{k}_2, \tag{6.9}$$

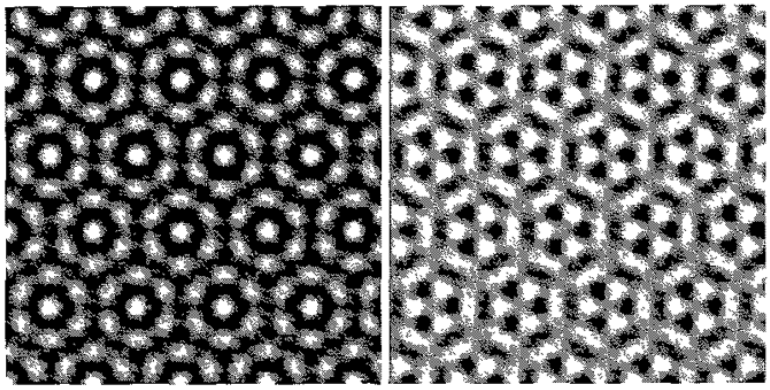
$$\boldsymbol{\kappa}_4 = \alpha\boldsymbol{k}_1 + (\alpha - \beta)\boldsymbol{k}_2, \tag{6.10}$$

$$\boldsymbol{\kappa}_5 = -\beta\boldsymbol{k}_1 - \alpha\boldsymbol{k}_2, \tag{6.11}$$

$$\boldsymbol{\kappa}_6 = (-\alpha + \beta)\boldsymbol{k}_1 + \beta\boldsymbol{k}_2, \tag{6.12}$$

where α and β are integers satisfying $\alpha > \beta > \alpha/2 > 0$. If α and β are relatively prime and $\alpha + \beta$ is not a multiple of 3 then these representations are translation-free. Figure 6.3 shows how the twelve-dimensional representation with $\alpha = 3$, $\beta = 2$ and $k_c = 7$ relates to the six-dimensional fundamental representation.

There are new axial solution branches in the twelve-dimensional representations, including super hexagons ($z_1 = z_2 = z_3 = z_4 = z_5 = z_6 \in \mathbb{R}$) and, when the $z \to -z$ reflection symmetry is present so that the symmetry group is $(D_6 \times \mathbb{Z}_2) \ltimes T^2$, super triangles ($z_1 = z_2 = z_3 = z_4 = z_5 = z_6 \in i\mathbb{R}$). Examples of super hexagons and super triangles are plotted in Figure 6.4. Again there are hidden symmetries in the problem, and this time they do restrict the form of the evolution equations for the z_j. Full details can be found in Dionne, Silber and Skeldon (1997).



(a) (b)

Fig 6.4. Greyscale plots of $u(x, y, t)$ as given in equation (6.6) for (a) super hexagons ($z_1 = z_2 = z_3 = z_4 = z_5 = z_6 = 1$) and (b) super triangles ($z_1 = z_2 = z_3 = z_4 = z_5 = z_6 = i$)

6.2 Mode interactions

Up to this point we have assumed that there is only one critical wavenumber k_c. What happens if that is not true? There is a multitude of possible alternative scenarios, but we want to stick to situations where the patterns will look regular and periodic, so let us look at resonant mode interactions. The simplest example is when we have modes $e^{ik_c x}$ and $e^{2ik_c x}$ contributing to a pattern in one spatial dimension ($x \in \mathbb{R}$), both arising at steady bifurcations from the zero solution under the influence of external stresses described by bifurcation parameters μ_1 and μ_2 respectively. Both modes fit on the same one-dimensional lattice whose basic period is $2\pi/k_c$. As usual we will set $k_c = 1$ by scaling the x variable. The solution can then be written in the form

$$u(x, t) = z_1(t)e^{ix} + z_2(t)e^{2ix} + c.c. + h.o.t., \quad z_1, z_2 \in \mathbb{C}. \tag{6.13}$$

We will assume that the system has $O(2)$ symmetry, and so we must have equivariance under

(i) translations: $(z_1, z_2) \to (z_1 e^{-ip}, z_2 e^{-2ip})$;
(ii) reflection: $(z_1, z_2) \to (\bar{z}_1\ \bar{z}_2)$

The e^{ix} and e^{2ix} modes individually correspond to different representations of $O(2)$. Mode interactions in general involve the interplay of different representations of the relevant symmetry group.

This problem has been investigated in detail by Proctor and Jones (1988) who were studying convection in two layers of fluid, one on top of the other with a thin heat-conducting plate between them. If the top layer is approximately half as thick

as the bottom layer then 2 : 1 resonance is seen We will look briefly here at some of the phenomena they found

Using translation equivariance we can deduce the form of the equations up to cubic order to be

$$\frac{dz_1}{dt} = \mu_1 z_1 + a\bar{z}_1 z_2 - b_1|z_1|^2 z_1 - b_2|z_2|^2 z_1, \tag{6.14}$$

$$\frac{dz_2}{dt} = \mu_2 z_2 + c z_1^2 - d_1|z_2|^2 z_2 - d_2|z_1|^2 z_2, \tag{6.15}$$

and using reflection equivariance we see that all coefficients must be real. If we rescale the amplitudes z_1 and z_2 we can set $a = 1$ and $c = \pm 1$. Now we will rewrite the equations by setting $z_1 = Re^{i\phi}$, $z_2 = Se^{i\psi}$ and separate out the real and imaginary parts of the equations to get

$$\frac{dR}{dt} = \mu_1 R + RS\cos(\psi - 2\phi) - b_1 R^3 - b_2 S^2 R, \tag{6.16}$$

$$R\frac{d\phi}{dt} = RS\sin(\psi - 2\phi), \tag{6.17}$$

$$\frac{dS}{dt} = \mu_2 S \pm R^2\cos(\psi - 2\phi) - d_1 S^3 - d_2 R^2 S, \tag{6.18}$$

$$S\frac{d\psi}{dt} = \mp R^2\sin(\psi - 2\phi) \tag{6.19}$$

Combining the second and fourth equations gives

$$\frac{d\chi}{dt} = \left(\mp\frac{R^2}{S} - 2S\right)\sin\chi, \tag{6.20}$$

where $\chi = \psi - 2\phi$, and so this system is in fact only three-dimensional, with dependent variables R, S and χ. The redundancy of the fourth equation is related to the freedom to choose the origin of x If we now write $X = S\cos\chi$, $Y = S\sin\chi$ and $Z = R^2$, the equations can be made to look neater:

$$\frac{dX}{dt} = \mu_2 X \pm Z + 2Y^2 - d_1 X(X^2 + Y^2) - d_2 XZ, \tag{6.21}$$

$$\frac{dY}{dt} = \mu_2 Y - 2XY - d_1 Y(X^2 + Y^2) - d_2 YZ, \tag{6.22}$$

$$\frac{dZ}{dt} = 2Z(\mu_1 + X - b_1 Z - b_2(X^2 + Y^2)). \tag{6.23}$$

The nontrivial fixed points are

(i) pure modes

$$Y = Z = 0,\ X^2 = \frac{\mu_2}{d_1}; \tag{6.24}$$

(ii) mixed modes

$$Y = 0, \tag{6.25}$$

$$0 = \mu_1 + X - b_1 Z - b_2 X^2, \tag{6.26}$$

$$0 = \mu_2 X \pm Z - d_1 X^3 - d_2 X Z; \tag{6.27}$$

(iii) (in the case $c = -1$ only) travelling waves

$$Z = 2(X^2 + Y^2) = \frac{2(2\mu_1 + \mu_2)}{(4b_1 + 2b_2 + 2d_2 + d_1)}, \tag{6.28}$$

$$X = \frac{\mu_2(b_2 + 2b_1) - \mu_1(d_1 + 2d_2)}{(4b_1 + 2b_2 + 2d_2 + d_1)} \tag{6.29}$$

Finding travelling waves at a steady bifurcation is quite surprising! They exist when $(2\mu_1 + \mu_2)(4b_1 + 2b_2 + 2d_2 + d_1) > 0$ and $|\kappa| \leq 1$, where $\kappa = X/\sqrt{X^2 + Y^2}$ The phases are then $\phi = \phi_0 + \omega t$ and $\psi = \psi_0 + 2\omega t$, where ϕ_0, ψ_0 and $\omega \equiv Y$ are constant The solution for $u(x, t)$ is

$$u(x, t) = \sqrt{Z} e^{i\phi_0} e^{i(x+\omega t)} + \sqrt{X^2 + Y^2} e^{i\psi_0} e^{2i(x+\omega t)} + c.c., \tag{6.30}$$

so you can see that this is indeed a travelling wave. The phase speed ω can be either negative, giving right-travelling waves, or positive, giving left-travelling waves The phase speed is zero at $Y = 0$, when $|\kappa| = 1$ or

$$[\mu_2(b_2 + 2b_1) - \mu_1(d_1 + 2d_2)]^2 = (2\mu_1 + \mu_2)(4b_1 + 2b_2 + 2d_2 + d_1). \tag{6.31}$$

The travelling waves are identical to the mixed modes at this point (the solution branches cross there) and bifurcate from them for $|\kappa| < 1$ Since the phase speed is zero at the point where the travelling waves bifurcate from the mixed modes, this is not a Hopf bifurcation It is a particularly unusual feature of this mode interaction that travelling waves bifurcate from a steady solution at a steady bifurcation This is a **drift instability** and is related to the fact that the phases ϕ and ψ of the mixed modes are arbitrary – only the combination $\chi = \psi - 2\phi$ is fixed. The drifting motion breaks the $x \to -x$ reflection symmetry of the mixed modes The mixed modes bifurcate in turn from the pure modes at

$$\mu_1 = \pm\sqrt{\frac{\mu_2}{d_1}} + \frac{b_2\mu_2}{d_1}, \tag{6.32}$$

where the two solution branches cross.

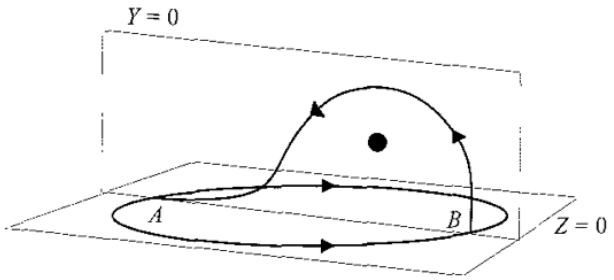


Fig 6.5 Heteroclinic orbit in the 2 : 1 mode interaction problem Point A is the pure mode $(-\sqrt{\mu_2/d_1}, 0, 0)$ and point B is $(\sqrt{\mu_2/d_1}, 0, 0)$ The black dot represents the mixed mode solution in the $Y = 0$ plane.

You can also find heteroclinic orbits in this system. For μ_1, μ_2 and d_1 all positive and b_2 such that the inequalities

$$\mu_1 + \sqrt{\frac{\mu_2}{d_1}} - \frac{b_2\mu_2}{d_1} > 0, \tag{6 33}$$

$$\mu_1 - \sqrt{\frac{\mu_2}{d_1}} - \frac{b_2\mu_2}{d_1} < 0, \tag{6 34}$$

hold, the trivial solution $X = Y = Z = 0$ is an unstable node, the pure mode, $X = \sqrt{\mu_2/d_1}$, $Y = Z = 0$, is stable in the X and Y directions and unstable in the Z direction, while the pure mode at $X = -\sqrt{\mu_2/d_1}$, $Y = Z = 0$, is stable in the X and Z directions and unstable in the Y direction In the invariant plane, $Z = 0$, these are the only fixed points, and we also have

$$\frac{\mathrm{d}}{\mathrm{d}t}(X^2 + Y^2) = \mu_2(X^2 + Y^2) - d_1(X^2 + Y^2)^2, \tag{6 35}$$

so the circle $X^2 + Y^2 = \mu_2/d_1$ is invariant. In fact the circle consists of two trajectories from $(-\sqrt{\mu_2/d_1}, 0, 0)$ to $(\sqrt{\mu_2/d_1}, 0, 0)$ as shown in Figure 6.5

Now the pure mode $(\sqrt{\mu_2/d_1}, 0, 0)$ is unstable to perturbations in the Z direction, so we will now look in the invariant plane $Y = 0$ to see what happens to trajectories leaving the fixed point The evolution equations are

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mu_2 X \pm Z - d_1 X^3 - d_2 XZ, \tag{6 36}$$

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = 2Z(\mu_1 + X - b_1 Z - b_2 X^2). \tag{6 37}$$

The nontrivial fixed points in this plane are the pure modes, $Z = 0$, $X = \pm\sqrt{\mu_2/d_1}$, and the mixed modes satisfying equations (6.26) and (6 27). We expect a trajectory joining $(\sqrt{\mu_2/d_1}, 0, 0)$, which is a saddle in the plane $Y = 0$, to $(-\sqrt{\mu_2/d_1}, 0, 0)$, which is a stable node, as long as the mixed mode solutions

are unstable nodes or foci. The mixed modes will be unstable in the full (X, Y, Z) space if the following inequalities hold:

$$2d_1X^2 + 2b_1Z \pm \frac{Z}{X} < 0, \tag{6.38}$$

$$b_1(\pm\frac{Z}{X} + 2d_1X^2) + (1 - 2b_2X)(\mp 1 + d_2X) > 0, \tag{6.39}$$

$$\mu_2 - 2X - d_2Z - d_1X^2 > 0. \tag{6.40}$$

In fact the mixed mode undergoes a Hopf bifurcation when the inequality is replaced by equality in (6.38). For small μ_1 and μ_2, the location of the Hopf bifurcation is given by

$$X \approx -\mu_1, \; Z \approx -2d_1\mu_1^3, \; \mu_2 \approx 3d_1\mu_1^2. \tag{6.41}$$

Inequality (6.39) must also hold for the Hopf bifurcation. Again for small μ_1 and μ_2 close to onset of the Hopf bifurcation, Z/X will be $O(\mu_1^2)$, and so we can see that (6.39) will hold if we take the second plus sign. It turns out, though it is difficult to show analytically, that the mixed modes can indeed be unstable in a region where μ_1, μ_2 and d_1 are all positive and b_2 satisfies inequalities (6.33) and (6.34). This means that there are heteroclinic orbits consisting of one of the half-circles shown in Figure 6.5, and a trajectory in the plane $Y = 0$ from $(\sqrt{\mu_2/d_1}, 0, 0)$ to $(-\sqrt{\mu_2/d_1}, 0, 0)$. Trajectories spiral out from the mixed-mode fixed point towards the heteroclinic orbit. The pure-mode fixed points are only unstable in the direction out along the heteroclinic orbit, so we expect the orbit to be stable. It also exists for a range of values of μ_1 and μ_2, so it is structurally stable too.

Of course, there is a huge number of possible mode interactions that we could investigate. The example we have just worked through is a steady-steady mode interaction, in that both modes were undergoing a steady bifurcation, but we could also have Hopf–Hopf or steady-Hopf interactions, interactions between more than two different modes, mode interactions on square, rhombic or hexagonal lattices, interactions between spatially varying and spatially uniform modes, and so on. The idea is always the same: write down the form of the solution $u(x, t)$ in terms of the contributing modes, and use equivariance to deduce the amplitude equations that you can use to investigate the system.

6.3 Spatial-period-multiplying bifurcations

What happens if, instead of starting with a spatially uniform solution, such as $u = 0$, we have a lattice pattern as the initial condition and want to look at bifurcations from that? This situation is actually quite easy to analyse if we assume

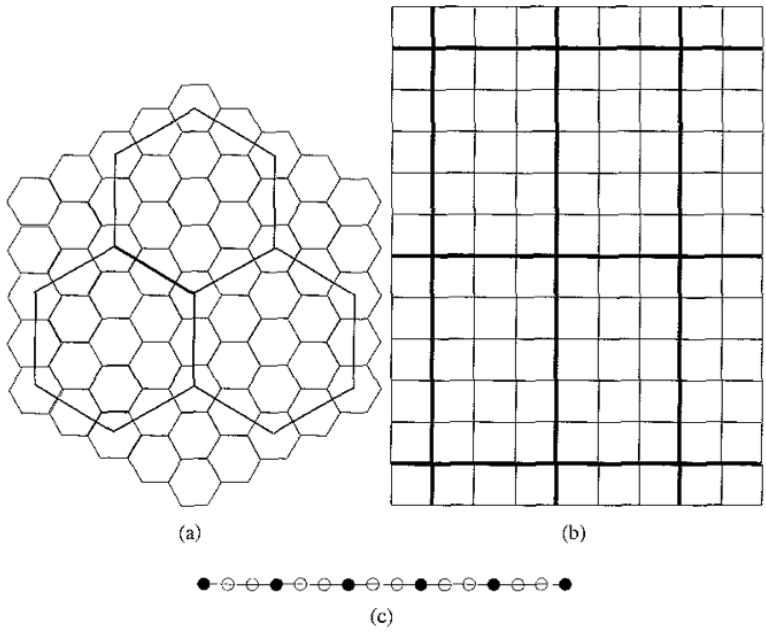
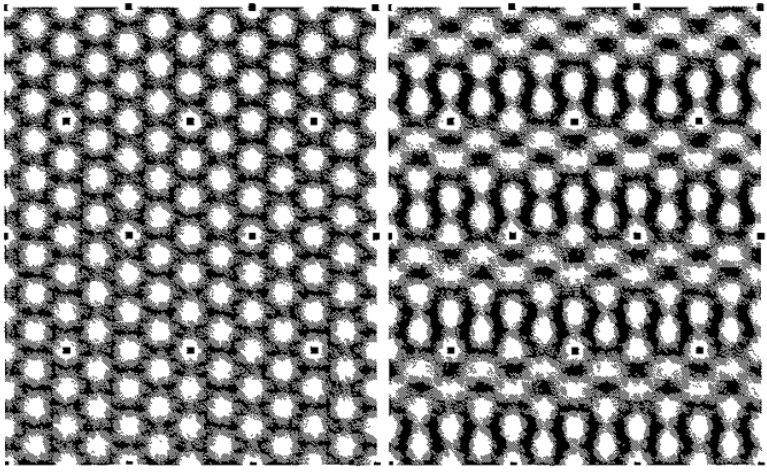


Fig. 6.6 Examples of superimposed lattices for spatial-period-multiplying bifurcations: (a) coarse and fine hexagonal lattices, (b) a coarse rectangular lattice over a fine square lattice and (c) coarse (filled circles) and fine (empty circles) one-dimensional lattices with periods in the ratio 3 : 1.

that the bifurcating solutions lie on a coarser lattice that fits on top of the original one: some examples are shown in Figure 6.6. Having set up the two lattices, it is straightforward to work out the symmetry group of the initial state on the coarse lattice, and apply equivariance as usual. The choice of coarse lattice will depend on the results you are trying to interpret with your analysis. For example, the coarse hexagonal lattice shown in Figure 6.6a was chosen by Tse *et al.* (2000) to interpret a Faraday experiment showing hexagons becoming unstable to a stripy modulation that was periodic on the coarse lattice. The basic hexagons and the modulated pattern are shown in Figure 6.7. The choice of lattice is not unique, and in the Faraday experiment example we could have chosen a rectangular lattice with sides in the ratio $\sqrt{3} : 1$, however it seemed natural to choose a larger hexagon as the basic unit of the lattice. In general it is good to choose the finest coarse lattice that the bifurcating pattern will fit onto in order to keep the calculations as simple as possible, but you could pick any coarse lattice that fits. When the ratio of the spatial periods of the coarse and fine lattices becomes very large, it is more appropriate to use the modulation formalism introduced in the next chapter to describe the evolution of the underlying fine-scale pattern.



(a) (b)

Fig. 6.7. Computer-generated images of the initial (a) and final (b) patterns for a spatial-period-multiplying bifurcation on the hexagonal lattice shown in Figure 6.6a. Reproduced from *Physica D* **146**, Tse, D P, Rucklidge, A M., Hoyle, R B. and Silber, M., Spatial period-multiplying instabilities of hexagonal Faraday waves, 367–387, copyright (2000), with permission from Elsevier.

Consider the simple example given by the two one-dimensional lattices in the ratio 3 : 1 shown in Figure 6.6c. We will analyse a bifurcation from a stripe pattern to stripes modulated with a spatial period 3 times as long as the original. Let us write the solution in the form

$$u(x,t) = u_0(x) + z(t)e^{ix} + c.c. + h.o.t., \quad z \in \mathbb{C} \tag{6.42}$$

where $u_0(x)$ is the initial stripe pattern, and $z(t)e^{ix}$ represents the bifurcating mode at 3 times the initial wavelength. The initial pattern has spatial period $2\pi/3$, one-third that of the bifurcating mode, so it has symmetry under discrete translations generated by $\tau : x \to x + 2\pi/3$. If we assume that $u_0(x)$ has a reflection symmetry $m_x : x \to -x$, then a possible initial state is $u_0(x) = e^{3ix} + c.c.$ Though it is useful to have a particular form of $u_0(x)$ in mind for illustrative purposes, it is not necessary for the analysis that follows: in fact any $u_0(x)$ with the same symmetries would do, and in a real experiment we would expect the initial state to be fully nonlinear. The symmetries of the original stripe pattern act on the amplitude of the bifurcating mode according to

$$\tau : z \to ze^{-2\pi i/3}, \tag{6.43}$$

$$m_x : z \to \bar{z}. \tag{6.44}$$

The evolution equation for z must be equivariant under these operations, so to cubic order it takes the form

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \mu z + \alpha \bar{z}^2 - \beta |z|^2 z, \tag{6.45}$$

Writing $z = Re^{i\phi}$ ($R > 0, 0 \leq \phi < 2\pi$) and separating out the real and imaginary parts of the equation gives

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \mu R + \alpha R^2 \cos 3\phi - \beta R^3, \tag{6.46}$$

$$R\frac{\mathrm{d}\phi}{\mathrm{d}t} = -\alpha R^2 \sin 3\phi \tag{6.47}$$

The coefficient α can be chosen to be positive without loss of generality, by changing the sign of z if necessary. We also choose β to be positive so that the cubic term is negative in the region $\mu > 0$ where the linear term is positive. In this case the cubic term is said to **saturate** or **quench** the linear instability, and acts to stabilise finite-amplitude solutions. This is typically more interesting than the case $\beta < 0$ when the cubic term is destabilising. The possible stationary solutions for positive α and β are:

(i) the trivial solution, $R = 0$;
(ii) modulated stripes (M_+)

$$\mu + \alpha R^2 - \beta R^3 = 0, \tag{6.48}$$

$$\phi = 0, 2\pi/3, 4\pi/3; \tag{6.49}$$

(iii) modulated stripes (M_-)

$$\mu - \alpha R^2 - \beta R^3 = 0 \tag{6.50}$$

$$\phi = \pi/3, \pi, 5\pi/3 \tag{6.51}$$

For $\mu > 0$, the trivial solution and M_- solutions are unstable, while the M_+ solution is stable for $R > \alpha/2\beta$. The bifurcation diagram is plotted in Figure 6.8. A modulated stripe solution is shown in Figure 6.9.

Maybe this has reminded you of the bifurcation on the hexagonal lattice. The similarity comes from the fact that the discrete translations generated by τ, when factored out by the coarse lattice, form the group $\mathbb{Z}_3$. The reflection m_x generates the group $\mathbb{Z}_2$, and together the translations and reflections generate $\mathcal{D}_3$, the symmetry group of an equilateral triangle. By writing the bifurcating mode in the form ze^{ix}, with $z \in \mathbb{C}$ we are using the two-dimensional natural representation of $\mathcal{D}_3$. The isotropy subgroups with one-dimensional fixed-point subspace are those generated by a reflection in the complex z plane, namely $\mathbb{Z}_2 = \{e, m\}$, where m is any one of $m_x\tau^n$ for $n = 0, 1, 2$. The fixed-point subspace of $m_x\tau^n$ is $Re^{ik\pi/3}$

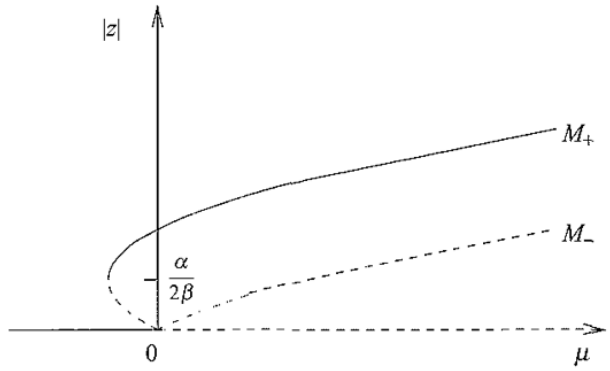


Fig. 6.8. Bifurcation diagram for the one-dimensional 3 : 1 spatial-period-multiplying bifurcation described by equation (6.45). The parameters α and β are both positive. The modulated stripe solutions M_+ and M_- are defined in equations (6.48)–(6.51). Solid lines represent stable solutions and dashed lines unstable solutions.

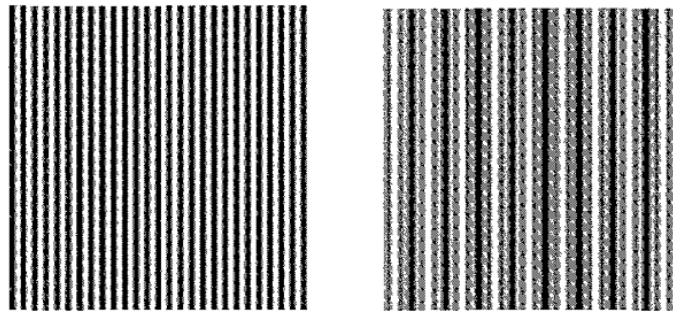


Fig. 6.9. (a) The initial stripe pattern and (b) the modulated stripe pattern M_+ with $\phi = 2\pi/3$ for the 3 : 1 spatial-period-multiplying bifurcation in one dimension.

with $k = 0$ or 3 when $n = 0$, $k = 2$ or 5 for $n = 1$ and $k = 1$ or 4 for $n = 2$, so the modulated stripe patterns that we found above are exactly the solutions predicted by the equivariant branching lemma.

6.4 Quasipatterns

Quasipatterns have local rotation and reflection symmetries, but lack spatial periodicity and so never repeat exactly in any direction. They look quite regular, as if they are generated by wavevectors spaced equally around the circle $|k| = k_c$, but not lying on a lattice. A dodecagonal (twelvefold) quasipattern seen in a Faraday wave experiment with two-frequency forcing by Arbell and Fineberg (2002) is

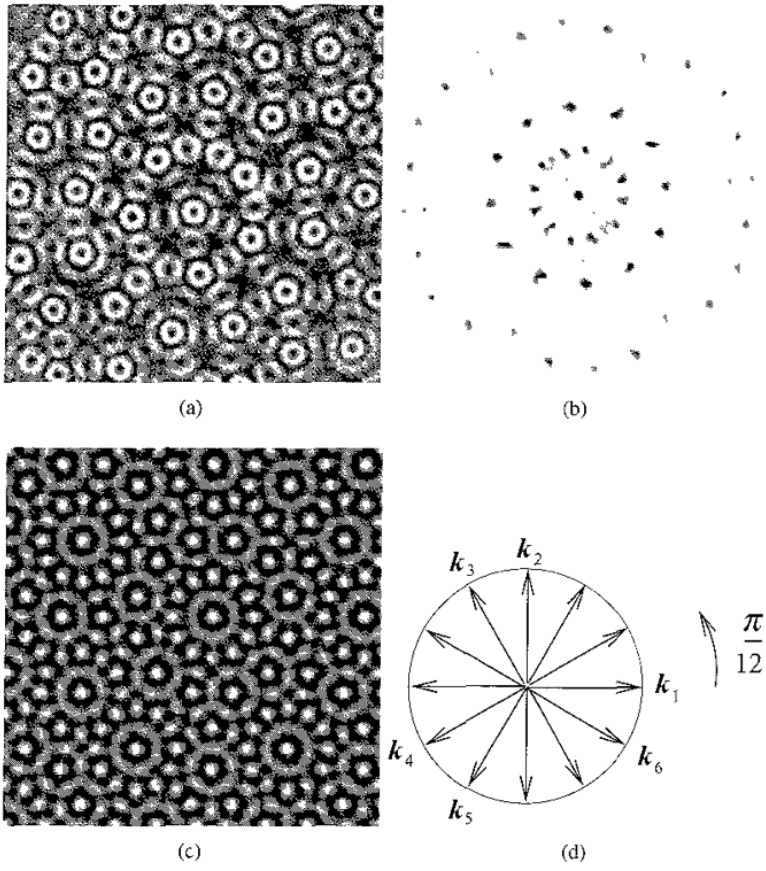


Fig. 6.10. (a) A twelvefold quasipattern seen in a Faraday wave experiment with two-frequency forcing, and (b) the corresponding spatial Fourier spectrum. (c) A greyscale image of the twelvefold quasipattern defined by equation (6.52) with $z_n = 1$ for all n and (d) the twelve wavevectors that generate it. Figures (a) and (b) are reproduced with permission from Arbell and Fineberg, *Physical Review E*, 65, 036224 (2002) © The American Physical Society (2002)

shown in Figure 6.10, together with a pattern generated from twelve wavevectors spaced equally around a circle according to

$$u(\boldsymbol{x}) = \sum_{n=1}^{6} z_n e^{i\boldsymbol{k}_n \cdot \boldsymbol{x}} + c.c., \tag{6.52}$$

where the $\boldsymbol{k}_n$ are given in Figure 6.10d, and in this case $z_n = 1$ for all n. There are two exact resonances in this set of wavevectors: $\boldsymbol{k}_1 + \boldsymbol{k}_3 + \boldsymbol{k}_5 = \boldsymbol{0}$ and $\boldsymbol{k}_2 + \boldsymbol{k}_4 + \boldsymbol{k}_6 = \boldsymbol{0}$, corresponding to two hexagonal sublattices.

If $\boldsymbol{x} \equiv (x_1, x_2) \in \mathbb{R}^2$ then the spatial frequency of modes $n = 4, 6$ in the x_1 direction is $\sqrt{3}/2$, which is not commensurate with the frequencies of modes

$n = 1, 3, 5$, which are either 1 or $1/2$. The set of spatial frequencies in the x_2 direction is the same, but taken in a different order, so the pattern is quasiperiodic in both the x_1 and x_2 directions. A quasipattern can be constructed from any set of wavevectors with incommensurate spatial frequencies in two directions. The more symmetric the set of wavevectors, the more regular the quasipattern will appear to be. The quasipatterns that are seen in experiments appear to be pretty regular. Edwards and Fauve (1993) found regular twelvefold quasipatterns in a Faraday wave experiment using a container the shape of France (Figure 1.2), demonstrating both an admirable appreciation of national geography, and the unimportance of boundary conditions for quasipattern formation. So the regularity seems to be intrinsic. While twelvefold quasipatterns seem to be the most common in planar systems, others, such as eight- and tenfold structures do occur (see, for example, Arbell and Fineberg, 2002). In fact, Arbell and Fineberg can tune their Faraday wave experiment to make $2n$-fold quasipatterns for any n. Three-dimensional quasicrystals are also possible: the first ones observed had icosahedral symmetry (Shechtman *et al.* 1984).

One immediate consequence of having incommensurate wavenumbers is that quasipatterns do not fit on a lattice. If you remember, the advantage of restricting bifurcation problems to a lattice was twofold: first only a finite (and small) number of wavevectors could lie on the critical circle so we could project the dynamics onto a small number of modes, reducing the complexity of the problem, and second, more importantly, the remaining modes were bounded away from the critical circle so we could apply the centre manifold theorem. Now the problem with quasipatterns is that by adding sufficiently many of the incommensurate wavevectors together in suitable combinations you can get as close to the critical circle as you want to (see Figure 6.11), so that even if there are still only a few critical modes, there is no guarantee that we can project the dynamics onto them. The combinations of modes coming close to the critical circle represent high-order near-resonances. A true resonance, such as occurs in the hexagonal lattice where we have $\boldsymbol{k}_1 = -\boldsymbol{k}_2 - \boldsymbol{k}_3$, so that the sum of $\boldsymbol{k}_2$ and $\boldsymbol{k}_3$ lies on the critical circle, leads to the appearance of additional terms in the amplitude equations – in the hexagonal case the $\bar{z}_2\bar{z}_3$ in the $\mathrm{d}z_1/\mathrm{d}t$ equation. If the near-resonance is very close to being exact then perhaps it makes no sense to ignore these extra terms. Another way of looking at the problem when amplitude equations are derived directly from the governing partial differential equations, as in the next chapter, is to say that these near-resonances produce small divisors in some coefficients in the amplitude equations, so these terms are much larger than expected and the expansion in $|z| \ll 1$ breaks down.

Despite concerns over their validity you can still write down amplitude equations for quasipatterns that are equivariant with respect to the symmetries of the set

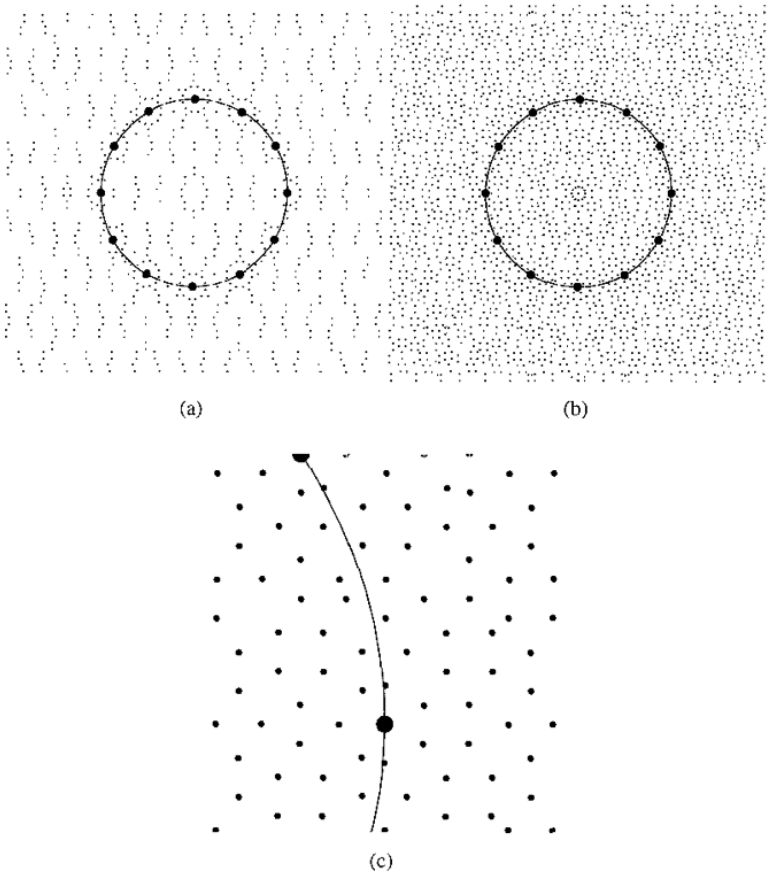


Fig. 6.11 Combinations of up to N of the twelve critical wavevectors of the dodecagonal quasipattern, defined by equation (6.52) approach the critical circle as N increases. The figures show the positions of all such combinations for (a) $N = 11$, (b) $N = 15$; (c) detail of (b). The critical circle is shown as a solid line, and the original twelve modes as large dots. The positions of integer combinations of the critical modes are shown as small dots. As N increases, the density of points increases and points get closer to the critical circle. Reproduced from Figure 5 of Rucklidge, A. M., Pattern formation in large domains, *Phil. Trans R Soc Lond A* **361**, 2649–64 (2003) with the permission of the Royal Society

of contributing wavevectors and to translations in two orthogonal directions. The problem, of course, is that these don't form a compact group. Using our twelvefold quasipattern, defined in equation (6.52), as an example, the relevant symmetries are translations $\boldsymbol{p} : z_n \to z_n e^{-i\boldsymbol{k}_n \cdot \boldsymbol{p}}$, and the symmetries of D_{12} generated by a rotation through $\pi/6$ ($\rho : z_1 \to z_6, z_2 \to \bar{z}_5, z_3 \to z_2, z_4 \to \bar{z}_1, z_5 \to z_4, z_6 \to \bar{z}_3$) and a reflection ($m : z_1 \to \bar{z}_1, z_3 \leftrightarrow \bar{z}_5, z_4 \leftrightarrow z_6$). If we now allow the pattern to vary in time, so that $z_n \equiv z_n(t)$, then to cubic order the amplitude equations that satisfy

these symmetries are

$$\frac{dz_i}{dt} = \mu z_i + \alpha \bar{z}_{i+2}\bar{z}_{i+4} - z_i\{\beta|z_i|^2 + \delta(|z_{i+2}|^2 + |z_{i+4}|^2) + \gamma(|z_{i-1}|^2 + |z_{i-3}|^2) + \nu|z_{i+1}|^2\}, \quad i = 1, 3, 5, \tag{6.53}$$

$$\frac{dz_i}{dt} = \mu z_i + \alpha \bar{z}_{i+2}\bar{z}_{i+4} - z_i\{\beta|z_i|^2 + \delta(|z_{i+2}|^2 + |z_{i+4}|^2) + \gamma(|z_{i+1}|^2 + |z_{i+3}|^2) + \nu|z_{i-1}|^2\}, \quad i = 2, 4, 6, \tag{6.54}$$

where the indices cycle with period 6, and where μ is the real bifurcation parameter and α, β, δ, γ and ν are real constants. Various authors have studied these equations or their equivalents (see, for example, Pismen, 1981, and Malomed, Nepomnyashchiĭ and Tribelskiĭ, 1989. More recently Echebarria and Riecke (2001) extended them to consider modulational phase instabilities, as we shall see in Chapter 9. The possible solutions are

(i) rolls, e.g. $Re(z_1) \neq 0$, $Im(z_1) = 0$, $z_2 = z_3 = z_4 = z_5 = z_6 = 0$;
(ii) rectangles, e.g. $Re(z_1) = Re(z_6) \neq 0$, $Im(z_1) = Im(z_6) = 0$, $z_2 = z_3 = z_4 = z_5 = 0$;
(iii) squares, e.g. $Re(z_1) = Re(z_2) \neq 0$, $Im(z_1) = Im(z_2) = 0$, $z_3 = z_4 = z_5 = z_6 = 0$;
(iv) hexagons, e.g. $Re(z_1) = Re(z_3) = Re(z_5) \neq 0$, $Im(z_1) = Im(z_3) = Im(z_5) = 0$, $z_2 = z_4 = z_6 = 0$;
(v) mixed modes, e.g. $0 \neq Re(z_1) = Re(z_3) \neq Re(z_5) \neq 0$, $Im(z_1) = Im(z_3) = Im(z_5) = 0$, $z_2 = z_4 = z_6 = 0$;
(vi) one-dimensional quasipatterns, e.g. $Re(z_1)Re(z_3)Re(z_5)Re(z_6) \neq 0$, $Im(z_1) = Im(z_3) = Im(z_5) = Im(z_6) = 0$, $z_2 = z_4 = 0$;
(vii) twelvefold quasipatterns, $Re(z_1) = Re(z_2) = Re(z_3) = Re(z_4) = Re(z_5) = Re(z_6) \neq 0$, $Im(z_1) = Im(z_2) = Im(z_3) = Im(z_4) = Im(z_5) = Im(z_6) = 0$.

The stability of the solutions can be determined using the amplitude equations. Typically there is hysteresis in the system, for example between the hexagon and dodecagonal (twelvefold) solutions (see Echebarria & Riecke, 2001, for further details).

If you write $z_n = R_n e^{i\phi_n}$, for $n = 1, \ldots, 6$, then the two global phases, $\Phi_1 = \phi_1 + \phi_3 + \phi_5$ and $\Phi_2 = \phi_2 + \phi_4 + \phi_6$, satisfy

$$\frac{d\Phi_j}{dt} = -\alpha R_j R_{j+2} R_{j+4} \left(\frac{1}{R_j^2} + \frac{1}{R_{j+2}^2} + \frac{1}{R_{j+4}^2} \right) \sin \Phi_j, \quad j = 1, 2 \tag{6.55}$$

The two phase modes

$$\phi_{x1} = \frac{1}{3}\left(\phi_1 - \frac{1}{2}\phi_3 - \frac{\sqrt{3}}{2}\phi_4 - \frac{1}{2}\phi_5 + \frac{\sqrt{3}}{2}\phi_6\right), \tag{6.56}$$

$$\phi_{x2} = \frac{1}{3}\left(\phi_2 + \frac{\sqrt{3}}{2}\phi_3 - \frac{1}{2}\phi_4 - \frac{\sqrt{3}}{2}\phi_5 - \frac{1}{2}\phi_6\right), \tag{6.57}$$

corresponding to translations in the x_1 and x_2 directions respectively, are undetermined as usual because there is freedom to choose the position of the origin. However, this still leaves undetermined two independent phase combinations, named **phason** modes,

$$\psi_1 = \frac{2}{3}\left(\phi_1 - \frac{1}{2}\phi_3 + \frac{\sqrt{3}}{2}\phi_4 - \frac{1}{2}\phi_5 - \frac{\sqrt{3}}{2}\phi_6\right), \tag{6.58}$$

$$\psi_2 = -\frac{2}{3}\left(\phi_2 - \frac{\sqrt{3}}{2}\phi_3 - \frac{1}{2}\phi_4 + \frac{\sqrt{3}}{2}\phi_5 - \frac{1}{2}\phi_6\right), \tag{6.59}$$

that correspond to translations of the two component hexagonal lattices relative to each other. This indeterminacy persists to all orders in the expansion of the amplitude equations, and reflects the fact that the twelve critical wavevectors do not form a lattice, and the symmetry group is noncompact. Strictly speaking the global phases are phason modes too, as this term is used for any phase modes that are not related to translations.

The upshot of all this is that you can write down amplitude equations for quasipatterns, and do some interesting analysis with them, but you can't be sure that the equations are valid or that your solutions approximate the quasipatterns that are observed in experiments. What is not in doubt is that quasipatterns are seen in the real world, and need explaining.

6.5 Pseudoscalar actions of $E(2)$

So far we have used the scalar action of the Euclidean group, $E(2)$, introduced in the previous chapter. However, $E(2)$ also has a **pseudoscalar** action defined by

$$\phi u(x) = \begin{cases} u(\phi^{-1}x), & \phi \in E(2) \text{ a translation or a rotation,} \\ -u(\phi^{-1}x), & \phi \in E(2) \text{ a reflection} \end{cases} \tag{6.60}$$

Most physically relevant Euclidean-invariant systems of partial differential equations reduce to a single equation that is equivariant with respect to the usual scalar action of $E(2)$, but there are some that reduce to a pseudoscalar partial differential equation, namely one that is equivariant under the pseudoscalar action defined

above (Melbourne, 1999). Notably, the two-dimensional Navier–Stokes equations for incompressible fluid flow in the plane reduce to a single pseudoscalar equation for the evolution of a streamfunction.

The pseudoscalar action leads to new planforms (Bosch Vivancos, Chossat & Melbourne, 1995; Bressloff *et al.*, 2001a), and though it turns out that physical fluid flows cannot exhibit them, it appears that they may be relevant to geometric visual hallucinations (see, for example, Bressloff *et al.*, 2001b). To see how the new planforms arise, consider the bifurcation on the hexagonal lattice, where to leading order the solutions are given by

$$u(x, y, t) = \sum_{j=1}^{3} z_j(t) e^{i\boldsymbol{k}_j \cdot \boldsymbol{x}} + c.c., \quad z_j \in \mathbb{C}, \tag{6.61}$$

with $\boldsymbol{k}_1 = (1, 0)$, $\boldsymbol{k}_2 = (-1/2, \sqrt{3}/2)$ and $\boldsymbol{k}_3 = (-1/2, -\sqrt{3}/2)$. As before, the relevant symmetry group is $D_6 \ltimes T^2$. The translations and rotations act just as in the scalar case:

(i) the rotation through $2\pi/3$, $\{\rho : (z_1, z_2, z_3) \to (z_3, z_1, z_2)\}$;
(ii) the rotation through π, $x \to -x$, $\{m_0 : (z_1, z_2, z_3) \to (\bar{z}_1, \bar{z}_2, \bar{z}_3)\}$;
(iii) translations $p \in T^2$ whose action is given by $\boldsymbol{p} \cdot (z_1, z_2, z_3) = (e^{-i\boldsymbol{k}_1 \cdot \boldsymbol{p}} z_1, e^{-i\boldsymbol{k}_2 \cdot \boldsymbol{p}} z_2, e^{-i\boldsymbol{k}_3 \cdot \boldsymbol{p}} z_3)$.

The difference comes with the reflections: the generating reflection in the vertical plane now acts as $\{m_v : (z_1, z_2, z_3) \to (-z_1, -z_3, -z_2)\}$, where the minus signs show that we are using the pseudoscalar action.

If we now try to deduce the amplitude equations, as we did in Section 5.4, we find that m_v leaves the term $a\bar{z}_2\bar{z}_3$ invariant, but transforms $\mathrm{d}z_1/\mathrm{d}t$ to $-\mathrm{d}z_1/\mathrm{d}t$, so the quadratic terms in the amplitude equations are forbidden under the pseudoscalar action of $E(2)$. The new equations are

$$\frac{\mathrm{d}z_1}{\mathrm{d}t} = \mu z_1 - b|z_1|^2 z_1 - c(|z_2|^2 + |z_3|^2) z_1, \tag{6.62}$$

$$\frac{\mathrm{d}z_2}{\mathrm{d}t} = \mu z_2 - b|z_2|^2 z_2 - c(|z_3|^2 + |z_1|^2) z_2, \tag{6.63}$$

$$\frac{\mathrm{d}z_3}{\mathrm{d}t} = \mu z_3 - b|z_3|^2 z_3 - c(|z_1|^2 + |z_2|^2) z_3, \tag{6.64}$$

which are the same as the cubic truncation of the amplitude equations for the bifurcation with midplane reflection in Section 5.4.1.

The most important differences with the scalar case are in the predicted planforms. Consider the solution

$$u_{\mathrm{ar}}(x_1, x_2) = a(e^{ix_1} + e^{-ix_1}), \tag{6.65}$$

where a is a real nonzero constant. In the scalar case this was a roll solution with isotropy subgroup $\mathcal{D}_2 \ltimes S^1$, whose elements are translations $(0, t) \in I^2$ in the x_2-direction, a rotation m_o through π, the reflection m_v ($x_2 \to -x_2$), and the reflection $m_o m_v$ ($x_1 \to -x_1$). Now, the reflections are no longer symmetries, since $m_v u_{ar} = m_o m_v u_{ar} = -u_{ar}$. However the combination of each of the reflections with a translation $x_1 \to x_1 + \pi$ are symmetries instead. The isotropy subgroup of u_{ar} is $\mathcal{D}_2^- \ltimes S^1$, where $\mathcal{D}_2^-$ is a **twisted** version of $\mathcal{D}_2$. The new planform u_{ar} is called **anti-rolls**. The leading order expression given in equation (6.65) is no different from that of a standard roll solution, but its interpretation in physical space leads to a planform with different symmetries. For example, if $u(x_1, x_2)$ represents a streamfunction, then a two-dimensional fluid flow is derived from it according to $v = -\nabla \times u\hat{x_3}$, where $\hat{x_3}$ is a unit vector in the x_3 direction. For u_{ar} this gives

$$v = (0, ai(e^{ix_1} - e^{-ix_1})). \tag{6.66}$$

Clearly the flow has the symmetries of $\mathcal{D}_2^- \ltimes S^1$, while u_{ar} does not. A flow with the symmetries of anti-rolls is illustrated in Figure 6.12a.

Similarly at leading order the standard hexagonal solution is replaced by the planform **simple oriented hexagons**. Both have the same leading-order expansion

$$u_{oh} = a(e^{ik_1 x} + e^{ik_2 x} + e^{ik_3 x}) + c.c., \quad a \in \mathbb{R},\ a \neq 0, \tag{6.67}$$

but simple oriented hexagons have isotropy subgroup $\mathbb{Z}_6$ generated by $m_o\rho$, a rotation through $\pi/3$. In this case the reflections cannot combine with other transformations to give symmetries of the pattern. The flow derived from this eigenfunction has the symmetries of the flow shown in Figure 6.12b.

Both anti-rolls and simple oriented hexagons are axial, as might be expected since they are the replacements for the axial roll and hexagonal patterns in the scalar case. In the pseudoscalar case there are two further axial planforms, namely **anti-triangles**, and **anti-rectangles**. Again this is perhaps not surprising, as the cubic truncation of the equations is the same as that for the bifurcation with midplane symmetry and scalar action, where similar solutions, regular triangles and the patchwork quilt are axial (see Table 5.3). The amplitude equations will however differ at higher order. The anti-triangles are given by

$$u_{tl} = ia(e^{ik_1 x} + e^{ik_2 x} + e^{ik_3 x}) + c.c., \quad a \in \mathbb{R},\ a \neq 0, \tag{6.68}$$

which is the same form as regular triangles take in the midplane scalar case, but now the isotropy subgroup is $\mathcal{D}_3$ generated by ρ and $m_o m_v : (z_1, z_2, z_3) \to (-\bar{z}_1, -\bar{z}_3, -\bar{z}_2)$. The anti-rectangle solution is

$$u_{pa} = a(e^{ik_2 x} - e^{ik_3 x}) + c.c., \quad a \in \mathbb{R},\ a \neq 0, \tag{6.69}$$

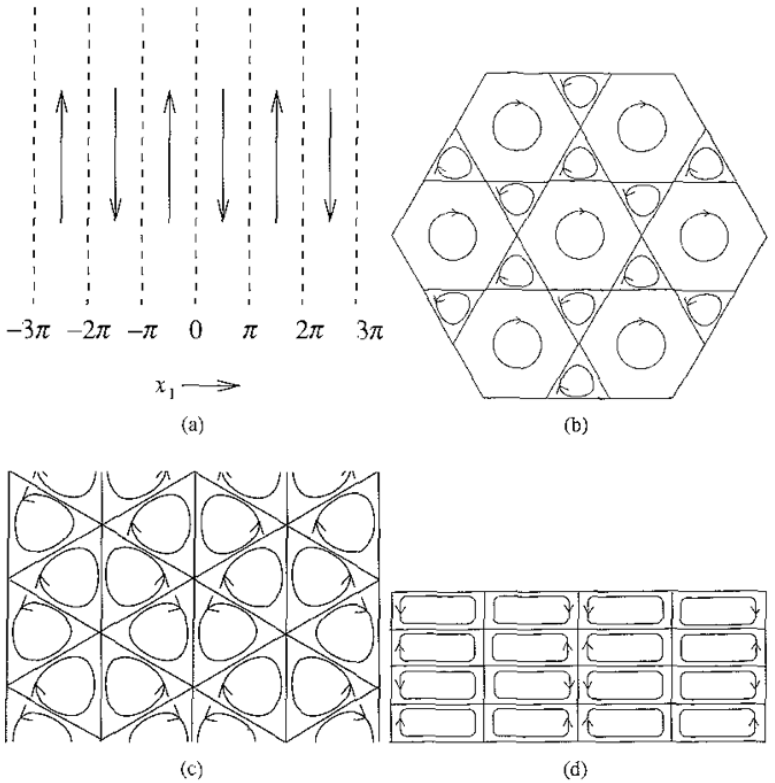


Fig 6 12. Examples of planforms with the symmetry of (a) anti-rolls (b) simple oriented hexagons, (c) anti-triangles and (d) anti-rectangles. The arrows indicate the direction of flow.

which has isotropy subgroup D_2, generated by m_v and m_0 as before. Examples of flows with the symmetries of anti-triangles and anti-rectangles are shown in Figures 6.12c and d.

The corresponding analysis on a square lattice shows that the pseudoscalar action leads to a new planform **simple anti-squares**, and on the rhombic lattice we find **anti-rhombs**

For the rest of this book we will be using the scalar action of $E(2)$ unless otherwise stated.

6.6 Hidden symmetries

It can happen that we choose to study a symmetric system in a context that restricts its symmetry in some way This might sound strange, but it can arise quite naturally: for example, we often want to consider systems where the governing equations are homogeneous and isotropic, and so have full Euclidean, $E(2)$, symmetry,

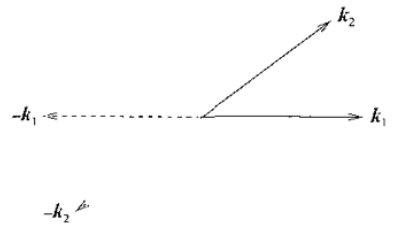


Fig. 6.13 The wavevectors k_1 and k_2 define a rhombic (dual) lattice. If the system is rotating, there is no reflection symmetry, and so there is no symmetry that maps either of the wavevectors to the other or its negative.

but where the pattern-forming experiment we want to analyse is being carried out in a finite domain with reduced symmetry, such as the square box we considered in Section 4.3.1. We expect then that the symmetry group of the bifurcation will be the symmetry group of the domain on which the boundary conditions are defined, in this case D_4, but the Euclidean symmetry of the governing equations and the form of the boundary conditions can sometimes give rise to additional hidden symmetries that we must take into account. These hidden symmetries are reflected in unexpected degeneracies of the amplitude equations in the restricted situation.

Hidden symmetries can also crop up if we restrict the bifurcation problem to a lattice in order to analyse periodic solutions. Some of the Euclidean symmetries that we 'ignore' in order to fit the problem onto a lattice turn out to lead to unexpected additional symmetries in the normal form equations.

Example 6.1 (Rotating convection on a rhombic lattice) *As a simple example, consider a two-dimensional system that is symmetric under all rotations and translations, a situation that arises naturally in convection in a rotating fluid layer (see Goldstein, Knobloch & Silber, 1990). The relevant symmetry group is* $SO(2) \ltimes \mathbb{R}^2$. *Now we will restrict the problem to a rhombic lattice, so that*

$$u(x, y, t) = z_1(t)e^{ik_1 x} + z_2(t)e^{ik_2 x} + c.c. + h.o.t., \tag{6.70}$$

where the angle, θ, *between* k_1 *and* k_2 *lies in the range* $0 < \theta < \pi/2$ *and is not equal to* $\pi/3$ *(to ensure that the lattice is neither square nor hexagonal). The wavevectors are shown in Figure 6.13. Using translation equivariance, the linear order amplitude equations are*

$$\frac{dz_1}{dt} = \mu_1 z_1, \tag{6.71}$$

$$\frac{dz_2}{dt} = \mu_2 z_2, \tag{6.72}$$

where μ_1 *and* μ_2 *are complex constants. Now since there is no reflection symmetry in the problem, and the only rotation that preserves the lattice is through* π, *there*

is no symmetry that relates the equation for z_1 to that for z_2, so it would seem that we can have $\mu_1 \neq \mu_2$. But this is nonsense! We know full well that the linear growth rates of the two modes must be equal, because the original system has full rotational symmetry, so the linear growth rate of a mode $e^{i\boldsymbol{k}\cdot\boldsymbol{x}}$ cannot depend on the direction of the wavevector, $\boldsymbol{k}$. From the point of view of the restricted problem on the rhombic lattice, the rotation symmetry is hidden, but it leads to the requirement that $\mu_1 = \mu_2$. This constraint is unexpected in the restricted context of the bifurcation on the rhombic lattice, and requires knowledge of the full system.

To formalise these ideas, imagine that we are once more dealing with a Γ-equivariant bifurcation problem

$$\frac{d\boldsymbol{x}}{dt} = f(\boldsymbol{x}, \mu), \quad \boldsymbol{x} \in V, \ \mu \in \mathbb{R}^n \tag{6.73}$$

where

$$\gamma f(\boldsymbol{x}, \mu) = f(\gamma \boldsymbol{x}, \mu), \quad \forall \gamma \in \Gamma \tag{6.74}$$

Now we reduce the symmetry of the problem by looking for fixed points of equation (6.73) that are invariant under a subgroup, $\Sigma \subsetneq \Gamma$, of the original symmetry group. Any such fixed point, $\boldsymbol{x}$, lies in $\mathrm{Fix}(\Sigma)$, since $\sigma \boldsymbol{x} = \boldsymbol{x}$ for all $\sigma \in \Sigma$. Fixed-point subspaces are flow-invariant, so we can restrict the bifurcation problem to $\mathrm{Fix}(\Sigma)$ and look for solutions there. Any restriction of the original problem that corresponds to a real experiment, such as conducting the experiment in a finite box, would lead naturally to the definition of a suitable flow-invariant subspace, since the dynamics of the equation cannot ignore the physical constraints imposed.

In the rotating convection example (Example 6.1), Γ would be $SO(2) \ltimes \mathbb{R}^2$, while Σ would be $\mathcal{L}$, where $\mathcal{L}$ is the group of translations that preserve the rhombic lattice. By restricting to functions that are periodic on the rhombic lattice, we are restricting to $\mathrm{Fix}(\mathcal{L})$.

6.6.1 Naive symmetries in the normalizer

$\mathrm{Fix}(\Sigma)$ is invariant under the normalizer $N(\Sigma)$, since

$$N(\Sigma) = \{\gamma \in \Gamma : \gamma^{-1}\Sigma\gamma = \Sigma\}, \tag{6.75}$$

and so if $\gamma \in N(\Sigma)$ and $\boldsymbol{x} \in \mathrm{Fix}(\Sigma)$ then

$$\sigma\gamma\boldsymbol{x} = \gamma\gamma^{-1}\sigma\gamma\boldsymbol{x} = \gamma\boldsymbol{x}, \ \forall \sigma \in \Sigma \tag{6.76}$$

since $\gamma^{-1}\sigma\gamma \in \Sigma$. Hence $\gamma\boldsymbol{x} \in \mathrm{Fix}(\Sigma)$ for all $\boldsymbol{x} \in \mathrm{Fix}(\Sigma)$ and $\gamma \in N(\Sigma)$. Thus $\gamma\mathrm{Fix}(\Sigma) \subseteq \mathrm{Fix}(\Sigma)$. Similarly $\gamma^{-1}\boldsymbol{x} \in \mathrm{Fix}(\Sigma)$ for all $\boldsymbol{x} \in \mathrm{Fix}(\Sigma)$ since $\gamma^{-1} \in$

$N(\Sigma)$, so $\mathrm{Fix}(\Sigma) \subseteq \gamma\mathrm{Fix}(\Sigma)$ and so we must have $\gamma\mathrm{Fix}(\Sigma) = \mathrm{Fix}(\Sigma)$. Furthermore if Σ is an isotropy subgroup Σ_x of a point x, then the normalizer is the largest group to leave $\mathrm{Fix}(\Sigma_x)$ invariant, since if there is a larger such group $\Delta \supsetneq N(\Sigma_x)$ then $\gamma x \in \mathrm{Fix}(\Sigma_x), \forall \gamma \in \Delta$, and so

$$\sigma\gamma x = \gamma x, \quad \forall \sigma \in \Sigma_x, \forall \gamma \in \Delta. \tag{6.77}$$

Hence we have

$$\gamma^{-1}\sigma\gamma x = x, \tag{6.78}$$

and so $\gamma^{-1}\sigma\gamma \in \Sigma_x$ for all $\sigma \in \Sigma_x$ and $\gamma \in \Delta$, and thus $\gamma^{-1}\Sigma_x\gamma \subseteq \Sigma_x$. Now if $\gamma^{-1}\Sigma_x\gamma \subsetneq \Sigma_x$ then there must be some $\sigma_1 \in \Sigma_x$ such that $\sigma_1 \neq \gamma^{-1}\sigma_2\gamma$, for any $\sigma_2 \in \Sigma_x$, but $\sigma_1 = \gamma^{-1}\gamma\sigma_1\gamma^{-1}\gamma$ and $\gamma\sigma_1\gamma^{-1} \in \Sigma_x$, since $\gamma^{-1} \in \Delta$, so there is a contradiction. Hence we must have $\gamma^{-1}\Sigma_x\gamma = \Sigma_x$, and so $\gamma \in N(\Sigma_x)$. So any element that maps $\mathrm{Fix}(\Sigma_x)$ into itself must lie in the normalizer.

The elements of the normalizer $N(\Sigma)$ map the restricted space of the problem, $\mathrm{Fix}(\Sigma)$, to itself and so are known as the **apparent** or **naive** symmetries in $\mathrm{Fix}(\Sigma)$ (Golubitsky, Marsden and Schaeffer, 1984). The group that we expect to govern the bifurcation in $\mathrm{Fix}(\Sigma)$ is $N(\Sigma)/\Sigma$, where we factor out Σ because it acts trivially on $\mathrm{Fix}(\Sigma)$.

Example 6.2 (Restriction to a planar lattice) *Now we can look at the restriction of a Euclidean-symmetric system to a planar lattice, $\mathcal{L}$, in a different way. Solutions that are periodic on the lattice lie in its fixed-point subspace,* $\mathrm{Fix}(\mathcal{L})$. *It is easy to check that the group of apparent symmetries, $N(\mathcal{L})$, is $H \ltimes \mathbb{R}^2$, where H is the holohedry of the lattice, and $\mathbb{R}^2$ is the group of two-dimensional translations. The group, $\mathcal{L}$, of discrete translations that preserve the lattice acts trivially on* $\mathrm{Fix}(\mathcal{L})$, *so we factor it out. This is equivalent to saying that we are only interested in what happens in one cell of the lattice, because we can reconstruct the others using periodicity. The resulting symmetry group is $H \ltimes T^2 \cong (H \ltimes \mathbb{R}^2)/\mathcal{L}$, and this governs the steady bifurcation on the lattice. We analysed the cases $H = \mathcal{D}_4$ and $H = \mathcal{D}_6$ in Sections 5.3 and 5.4.*

6.6.2 Hidden symmetries in the complement of the normalizer

Having decided what the 'obvious' symmetries are, we can now define the hidden symmetries. Naturally, these will be the ones that do not lie in the normalizer, but that nonetheless have the potential to affect the form of $f|_{\mathrm{Fix}(\Sigma)}$. One definition of a **hidden symmetry** is an element, $\gamma \in \Gamma$, that does not leave $\mathrm{Fix}(\Sigma)$ invariant, and so does not look like a symmetry of the restricted problem, but that satisfies

$$\gamma\mathrm{Fix}(\Sigma) \cap \mathrm{Fix}(\Sigma) \neq \{\mathbf{0}\} \tag{6.79}$$

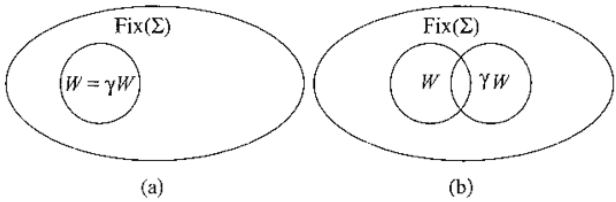


Fig 6 14. Hidden symmetries map a subspace $W \subset \mathrm{Fix}(\Sigma)$ (a) to itself or (b) to another subspace $\gamma W \subset \mathrm{Fix}(\Sigma)$ Case (b) is that of a subtle symmetry

(Manoel & Stewart, 2000). There are other possible definitions, as we shall see shortly Equation (6.79) can be satisfied either by mapping a subspace of $\mathrm{Fix}(\Sigma)$ to itself, or by mapping a subspace $W \subset \mathrm{Fix}(\Sigma)$ to a different subspace $\gamma W \subset \mathrm{Fix}(\Sigma)$ Golubitsky, Marsden and Schaeffer (1984) referred to γ as a **subtle** symmetry in the latter case. This term is not used any more, but it is worth bearing in mind that there need not necessarily be a subspace of $\mathrm{Fix}(\Sigma)$ on which γ acts as a symmetry in order for γ to be a hidden symmetry of the problem. The difference between the two situations is illustrated in Figure 6.14. In example 6 1 above, the rotation through θ that maps one wavevector of the rhombic lattice to the other would be a subtle symmetry, as it maps the subspace $(z_1, 0) \in \mathrm{Fix}(\mathcal{L})$ to the subspace $(0, z_2) \in \mathrm{Fix}(\mathcal{L})$ that overlaps only at $(0, 0)$.

Condition (6.79) says that there is a nonzero $\boldsymbol{x} \in \mathrm{Fix}(\Sigma)$ such that $\gamma\boldsymbol{x}$ also lies in $\mathrm{Fix}(\Sigma)$. Since $\mathrm{Fix}(\Sigma)$ is flow-invariant we will also have $f(\boldsymbol{x}, \mu) \in \mathrm{Fix}(\Sigma)$ and $f(\gamma\boldsymbol{x}, \mu) \in \mathrm{Fix}(\Sigma)$, so that $f(\boldsymbol{x}, \mu) = f|_{\mathrm{Fix}(\Sigma)}(\boldsymbol{x}, \mu)$ and $f(\gamma\boldsymbol{x}, \mu) = f|_{\mathrm{Fix}(\Sigma)}(\gamma\boldsymbol{x}, \mu)$ Then from the equivariance condition (6.74) on f we must have

$$\gamma f|_{\mathrm{Fix}(\Sigma)}(\boldsymbol{x}, \mu) = f|_{\mathrm{Fix}(\Sigma)}(\gamma\boldsymbol{x}, \mu) \tag{6.80}$$

for all $\boldsymbol{x} \in \gamma\mathrm{Fix}(\Sigma) \cap \mathrm{Fix}(\Sigma)$ Unexpectedly then, the hidden symmetry, γ, has the potential to restrict the form of $f|_{\mathrm{Fix}(\Sigma)}$. Not every hidden symmetry imposes a restriction in practice, since for a given γ, it may turn out that equation (6.80) is automatically satisfied for all f that are equivariant under the group of apparent symmetries, $N(\Sigma)$

Even if there are no hidden symmetries satisfying condition (6 79), equivariance of f on the full space V may still impose restrictions on the form of $f|_{\mathrm{Fix}(\Sigma)}$, since for $\boldsymbol{x} \in \mathrm{Fix}(\Sigma)$ we have

$$\gamma f|_{\mathrm{Fix}(\Sigma)}(\boldsymbol{x}, \mu) = \gamma f(\boldsymbol{x}, \mu) = f(\gamma\boldsymbol{x}, \mu), \quad \forall \gamma \in \Gamma. \tag{6.81}$$

To work out exactly what these restrictions are, we can no longer work only in the space $\mathrm{Fix}(\Sigma)$, because now $f(\gamma\boldsymbol{x}, \mu)$ does not lie there, so we must consider the whole space V Since any group element in $\Gamma - N(\Sigma)$, the complement of the normalizer, might restrict the form of $f|_{\mathrm{Fix}(\Sigma)}$ according to equation (6 81), there is

a case for defining all $\gamma \in \Gamma - N(\Sigma)$ to be hidden symmetries, while emphasising that not every hidden symmetry leads to unexpected effects In this philosophy there are four types of hidden symmetry, $\gamma \in \Gamma - N(\Sigma)$:

(i) γ satisfies equation (6 79) and maps $W \subset \mathrm{Fix}(\Sigma)$ to itself;
(ii) γ satisfies equation (6 79) and maps $W \subset \mathrm{Fix}(\Sigma)$ to a different subspace $\gamma W \subset \mathrm{Fix}(\Sigma)$;
(iii) γ does not satisfy equation (6 79), but restricts the form of $f|_{\mathrm{Fix}(\Sigma)}$;
(iv) γ neither satisfies equation (6 79) nor restricts the form of $f|_{\mathrm{Fix}(\Sigma)}$

Hidden symmetries that map a subspace of $\mathrm{Fix}(\Sigma)$ to itself can often be detected from within the restricted problem, even if we are ignorant of the extended group, Γ We may simply notice that there are additional symmetries in certain subspaces, which should alert us to the possibility that hidden symmetries are at work. However to find any remaining hidden symmetries we must know what Γ is Now if we have a problem with a hidden symmetry, then it may well be because we started out working in the restricted space, and had no idea that our problem was embedded in a bigger space with a bigger group acting on it until we came across some unexpected degeneracies in the equations or extra symmetries in the solutions. So having to know the right extended group in advance is very inconvenient! On the other hand, understanding that such hidden symmetries can exist can allow experimenters to modify their apparatus so as to eliminate them. We shall see this shortly in Section 6.7 on Neumann boundary conditions

Example 6.3 (2 : 1 mode interaction with $O(2)$ symmetry) *Imagine we have a bifurcation problem with $O(2)$ symmetry on the domain $-\pi \leq x \leq \pi$ with periodic boundary conditions, where $O(2)$ is generated by a translation $p : x \to x + x_0$ and a reflection $r : x \to -x$ Typically we can write solutions to a problem like this in the form*

$$u(x,t) = \sum_k a_k(t) e^{ikx} + c.c. + h.o.t\ , \tag{6 82}$$

where the wavenumbers k are integers and the amplitudes $a_k(t)$ are complex In general, solutions involving only one wavenumber k will bifurcate from the zero solution at the critical value of the bifurcation parameter, but under certain circumstances two different modes can bifurcate at once Consider the case where the $k = 1$ and $k = 2$ modes become unstable simultaneously We can write the solution as

$$u(x,t) = a_1 e^{ix} + a_2 e^{2ix} + \bar{a}_1 e^{-ix} + \bar{a}_2 e^{-2ix} + h.o\,t \tag{6 83}$$

Now in many applications we would only be interested in solutions that satisfy some further restriction, such as reflection symmetry. In this case the solutions

of interest would lie in $\mathrm{Fix}(\mathbb{Z}_2)$, *where* $\mathbb{Z}_2$ *is the reflection group generated by* $r : (a_1, a_2) \to (\bar{a}_1, \bar{a}_2)$ *The fixed-point subspace,* $\mathrm{Fix}(\mathbb{Z}_2)$, *is given by* $(a_1, a_2) = (A, B)$, *where* $A(t)$ *and* $B(t)$ *are real The naive symmetries in this problem are the elements of* $N(\mathbb{Z}_2)$, *which are generated by the reflection* r *and the translation by half a period,* $p_{1/2} : x \to x + \pi$, *which maps* (A, B) *to* $(-A, B)$, *so that* $\mathrm{Fix}(N(\mathbb{Z}_2)) = (0, B)$ *All the remaining translations are hidden symmetries in the problem*

The expected symmetry group of the bifurcation in $\mathrm{Fix}(\mathbb{Z}_2)$ *is* $N(\mathbb{Z}_2)/\mathbb{Z}_2$, *which is generated by* $p_{1/2}$. *The evolution equations for* $(A, B) \in \mathrm{Fix}(\mathbb{Z}_2)$ *must therefore be equivariant with respect to* $p_{1/2}$, *and so to cubic order we expect them to be*

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \mu_1 A + \alpha_1 \bar{A} B + \beta_1 A^3 + \beta_2 B^2 A, \tag{6.84}$$

$$\frac{\mathrm{d}B}{\mathrm{d}t} = \mu_2 B + \alpha_2 A^2 + \alpha_3 B^2 + \beta_3 A^2 B + \beta_4 B^3, \tag{6.85}$$

where all the coefficients μ_i, α_i *and* β_i *are real*

However, if we take the full problem and impose equivariance under the reflection, r, *and* **all** *translations,* p, *then to cubic order we find*

$$\frac{\mathrm{d}a_1}{\mathrm{d}t} = \widehat{\mu}_1 a_1 + \widehat{\alpha}_1 \bar{a}_1 a_2 + \widehat{\beta}_1 |a_1|^2 a_1 + \widehat{\beta}_2 |a_2|^2 a_1, \tag{6.86}$$

$$\frac{\mathrm{d}a_2}{\mathrm{d}t} = \widehat{\mu}_2 a_2 + \widehat{\alpha}_2 a_1^2 + \widehat{\beta}_3 |a_1|^2 a_2 + \widehat{\beta}_4 |a_2|^2 a_2, \tag{6.87}$$

where all the coefficients $\widehat{\mu}_i$, $\widehat{\alpha}_i$ *and* $\widehat{\beta}_i$ *are real. Restricting to* $\mathrm{Fix}(\mathbb{Z}_2)$, *where the amplitudes are real, produces the equations*

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \widehat{\mu}_1 A + \widehat{\alpha}_1 \bar{A} B + \widehat{\beta}_1 A^3 + \widehat{\beta}_2 B^2 A, \tag{6.88}$$

$$\frac{\mathrm{d}B}{\mathrm{d}t} = \widehat{\mu}_2 B + \widehat{\alpha}_2 A^2 + \widehat{\beta}_3 A^2 B + \widehat{\beta}_4 B^3 \tag{6.89}$$

Comparing these with equations (6.84) and (6.85) that we found by working naively in $\mathrm{Fix}(\mathbb{Z}_2)$, *we see that the hidden translation terms have forbidden the appearance of the* B^2 *term in the equation for* $\mathrm{d}B/\mathrm{d}t$.

It turns out that one of the hidden translations is a symmetry on a subspace of $\mathrm{Fix}(\mathbb{Z}_2)$ *The translation by a quarter period,* $p_{1/4} : x \to x + \pi/2$, $(0, B) \to (0, -B)$, *maps* $\mathrm{Fix}(N(\mathbb{Z}_2))$ *to itself, but does not leave* $\mathrm{Fix}(\mathbb{Z}_2)$ *invariant (since* $p_{1/4} \notin N(\mathbb{Z}_2)$*), and so condition (6.79) is satisfied with* $\gamma = p_{1/4}$ *and* $\Sigma = \mathbb{Z}_2$. *Even if we hadn't known that all translations were hidden symmetries, we might have guessed that* $p_{1/4}$ *was one if we had noticed that it mapped the subspace* $(0, B)$ *to itself.*

Now, on $\mathrm{Fix}(N(\mathbb{Z}_2)) = (0, B)$ *equations (6.84) and (6.85) reduce to the single equation*

$$\frac{\mathrm{d}B}{\mathrm{d}t} = \mu_2 B + \alpha_3 B^2 + \beta_4 B^3, \tag{6.90}$$

that must be equivariant with respect to $p_{1/4}$ *in order to satisfy equation (6.80). This is enough to show that* α_3 *must be zero.*

Golubitsky, Marsden and Schaeffer used this example in their 1984 paper where they coined the term 'hidden symmetry'.

Example 6.4 *An example where condition (6.79) is not satisfied, and yet the form of the evolution equation on the fixed-point subspace is more degenerate than would be expected, is given by the action of* $\mathcal{D}_5$ *on* $\mathbb{C}_2$. *The generators of the group are* ρ, *a rotation through* $2\pi/5$, $z \to ze^{-2\pi i/5}$, *and* m, *a reflection in the real axis* $z \to \bar{z}$. *Under these symmetries* $|z|^2$ *and* $Re(z^5)$ *are invariant, so we expect an evolution equation of the form*

$$\frac{\mathrm{d}z}{\mathrm{d}t} = g(|z|^2, Re(z^5))z + h(|z|^2, Re(z^5))\bar{z}^4, \tag{6.91}$$

where g *and* h *are real, smooth functions. Now consider restricting to* $\mathrm{Fix}(\mathbb{Z}_2) = \mathbb{R}$, *where* $\mathbb{Z}_2 = \{e, m\}$. *The normaliser* $N(\mathbb{Z}_2)$ *is just* $\mathbb{Z}_2$ *itself, and so acts trivially on* $\mathrm{Fix}(\mathbb{Z}_2)$. *There is no symmetry* $\gamma \in D_5 - \mathbb{Z}_2$ *that satisfies the condition (6.79), because* $\rho^n \mathrm{Fix}(\mathbb{Z}_2) \cap \mathrm{Fix}(\mathbb{Z}_2) = \{0\}$ *for* $n = 1, 2, 3, 4$. *Working naively in* $\mathrm{Fix}(\mathbb{Z}_2)$ *there would appear to be no restriction on the evolution equation for* $x \in \mathbb{R}$, *and so using a Taylor expansion about* $x = 0$ *we would write down the equation*

$$\frac{\mathrm{d}x}{\mathrm{d}t} = ax + bx^2 + \cdots \tag{6.92}$$

where a *and* b *are real constants and where we assume that there is a fixed-point at* $x = 0$ *so that there is no constant term on the righthand side. However, restricting equation (6.91) to the real line gives*

$$\frac{\mathrm{d}x}{\mathrm{d}t} = g(x^2, x^5)x + h(x^2, x^5)x^4, \tag{6.93}$$

which has no term in x^2 *on the righthand side. The symmetry that forces* b *on the righthand side of equation (6.92) to be zero is the hidden rotation,* ρ. *There is no quadratic term in* z, $\tilde{g}(z)$, *that is equivariant under* ρ, *so as to satisfy equation (6.81),*

$$e^{-2\pi i/5}\tilde{g}(x) = \tilde{g}(xe^{-2\pi i/5}). \tag{6.94}$$

Thus b *must be zero.*

An alternative perspective on this example, stressing the importance of smoothness for the extension of equivariants on $\mathrm{Fix}(\Sigma)$ *to* V *can be found in Manoel and Stewart (2000).*

6.6.3 Hidden symmetries and maximal isotropy subgroups

It is straightforward to show that $\gamma\mathrm{Fix}(\Sigma)$ is the fixed-point subspace of the subgroup $\gamma\Sigma\gamma^{-1}$ of Γ, since if $\boldsymbol{y} = \gamma\boldsymbol{x}$, with $\boldsymbol{x} \in \mathrm{Fix}(\Sigma)$, then

$$\gamma\sigma\gamma^{-1}\boldsymbol{y} = \gamma\sigma\gamma^{-1}\gamma\boldsymbol{x} = \gamma\sigma\boldsymbol{x} = \gamma\boldsymbol{x} = \boldsymbol{y}, \quad \forall\sigma \in \Sigma, \tag{6.95}$$

and hence $\boldsymbol{y} \in \mathrm{Fix}(\gamma\Sigma\gamma^{-1})$. Conversely, suppose $\boldsymbol{y} \in \mathrm{Fix}(\gamma\Sigma\gamma^{-1})$, then we have

$$\gamma\sigma\gamma^{-1}\boldsymbol{y} = \boldsymbol{y}, \quad \forall\sigma \in \Sigma, \tag{6.96}$$

and hence

$$\sigma\gamma^{-1}\boldsymbol{y} = \gamma^{-1}\boldsymbol{y}, \quad \forall\sigma \in \Sigma, \tag{6.97}$$

so that $\gamma^{-1}\boldsymbol{y} \in \mathrm{Fix}(\Sigma)$, and hence $\boldsymbol{y} \in \gamma\mathrm{Fix}(\Sigma)$. Putting the forwards and backwards implications together we have $\mathrm{Fix}(\gamma\Sigma\gamma^{-1}) = \gamma\mathrm{Fix}(\Sigma)$

Now if $\boldsymbol{x} \in \gamma\mathrm{Fix}(\Sigma) \cap \mathrm{Fix}(\Sigma)$, then

$$\delta\boldsymbol{x} = \boldsymbol{x}, \quad \forall\delta \in \Sigma \cup \gamma\Sigma\gamma^{-1}, \tag{6.98}$$

and so $\boldsymbol{x}$ lies in the fixed-point subspace of the subgroup Δ, the smallest subgroup containing all the elements of Σ and $\gamma\Sigma\gamma^{-1}$ The converse is also true, so

$$\gamma\mathrm{Fix}(\Sigma) \cap \mathrm{Fix}(\Sigma) = \mathrm{Fix}(\Delta). \tag{6.99}$$

If γ is a hidden symmetry satisfying condition (6.79), there must be some nonzero $\boldsymbol{x} \in \mathrm{Fix}(\Delta)$, with isotropy subgroup $\Delta_{\boldsymbol{x}} \supseteq \Delta$ We also have $\boldsymbol{x} \in \mathrm{Fix}(\Sigma)$ and hence $\Sigma \subseteq \Delta_{\boldsymbol{x}}$ If Σ is itself a maximal isotropy subgroup, we must have $\Sigma = \Delta_{\boldsymbol{x}}$, and therefore $\Sigma \supseteq \Delta$ and $\mathrm{Fix}(\Sigma) \subseteq \mathrm{Fix}(\Delta)$. Now from equation (6.99) it is clear that $\mathrm{Fix}(\Delta) \subseteq \mathrm{Fix}(\Sigma)$, but since $\gamma \notin N(\Sigma)$ we cannot have $\mathrm{Fix}(\Delta) = \mathrm{Fix}(\Sigma)$, and so there is a contradiction From this we conclude that when Σ is a maximal isotropy subgroup of Γ there are no hidden symmetries satisfying condition (6.79).

6.7 Hidden symmetries and reflecting boundary conditions

The idea of hidden symmetries will become clearer if we look at some more examples. We will start with a partial differential equation, equivariant under the

Euclidean group, $E(2)$:

$$\frac{\partial u(x, y, t)}{\partial t} = f(u(x, y, t), \mu), \tag{6 100}$$

where $x, y, t, \mu \in \mathbb{R}$ and where f is a nonlinear operator that may include spatial derivatives and that satisfies the equivariance condition

$$\gamma f(u, \mu) = f(\gamma u, \mu), \quad \forall \gamma \in E(2) \tag{6.101}$$

In Chapter 7 we will look in more detail at the way group elements act on a spatial operator, but for now it is enough to know that the equivariance condition takes the form we have been using throughout this chapter for ordinary differential equations

We will assume that equation (6.100) has a solution $u = 0$, and that this zero solution undergoes a bifurcation as μ is increased through zero. The solution $u = 0$ has full Euclidean symmetry, because it is the same everywhere and in every direction. If we were to study this bifurcation problem in the infinite plane, its symmetry group would be $E(2)$, but suppose that we are doing a real experiment and the experimental domain is the rectangle $0 \leq x \leq a$, $0 \leq y \leq b$. At first glance we would assume that the symmetry group of the problem is the symmetry group of the rectangular domain, generated by the reflections $m_x : x \to a - x$ and $m_y : b \to b - y$ in the lines $x = a/2$ and $y = b/2$ respectively, but there is something we have forgotten until now, namely the effect of the boundary conditions acting along the edges of the domain. Suppose our problem has reflecting boundary conditions such that

$$\frac{\partial^p u}{\partial x^p} = 0, \quad \text{on } x = 0 \text{ and } x = a, \quad \text{for } p \text{ odd} \tag{6 102}$$

$$\frac{\partial^q u}{\partial y^q} = 0, \quad \text{on } y = 0 \text{ and } y = b, \quad \text{for } q \text{ odd}. \tag{6.103}$$

Reflecting boundary conditions are often natural: for example, if u measures the concentration of some chemical in a reaction-diffusion system, then the Neumann boundary conditions

$$\frac{\partial u}{\partial x} = 0, \quad \text{on } x = 0 \text{ and } x = a, \tag{6 104}$$

$$\frac{\partial u}{\partial y} = 0, \quad \text{on } y = 0 \text{ and } y = b, \tag{6.105}$$

express the requirement that there is no flux of the chemical across the boundaries. Typically, for a reaction-diffusion system, the governing equation (6 100) will take

the form

$$\frac{\partial u}{\partial t} = D_x \frac{\partial^2 u}{\partial x^2} + D_y \frac{\partial^2 u}{\partial y^2} + g(u, \mu), \tag{6.106}$$

where D_x and D_y are real constants, and $g(u, \mu)$ is a nonlinear function of u that does not involve spatial derivatives. Differentiating equation (6.106) with respect to x and applying the Neumann boundary conditions shows that $\partial^3 u/\partial x^3 = 0$ on $x = 0$ and $x = a$. Differentiating twice more shows that $\partial^5 u/\partial x^5 = 0$ there too, and so on. The odd y-derivatives can be shown to vanish on the appropriate boundaries in a similar manner. The connection between Neumann boundary conditions and hidden symmetries in reaction-diffusion systems has been studied by Crawford *et al* (1991) and Gomes, Labouriau and Pinho (1999). In the rest of this section we will investigate some of their results in detail.

Any stationary solution, $u(x, y)$, of equation (6.100), satisfying the boundary conditions (6.102) and (6.103) can be extended to the region $-a \leq x \leq a$, $-b \leq y \leq b$ by reflecting in the boundaries such that

$$u(-x, y) = u(x, y), \tag{6.107}$$

$$u(x, -y) = u(x, y). \tag{6.108}$$

The new solution satisfies the reflecting boundary conditions on the original boundaries, and periodic boundary conditions on the extended domain.

Conversely, any solution $u(x, y)$ that satisfies periodic boundary conditions on the extended domain, and also the reflection conditions (6.107) and (6.108), will satisfy reflecting boundary conditions on the original domain, since differentiating the reflection conditions with respect to x and y respectively gives

$$-\frac{\partial^p u}{\partial x^p}(-x, y) = \frac{\partial^p u}{\partial x^p}(x, y), \quad \text{for } p \text{ odd}, \tag{6.109}$$

$$-\frac{\partial^q u}{\partial y^q}(x, -y) = \frac{\partial^q u}{\partial y^q}(x, y), \quad \text{for } q \text{ odd}, \tag{6.110}$$

$$\tag{6.111}$$

and so we immediately have

$$-\frac{\partial^p u}{\partial x^p}(0, y) = \frac{\partial^p u}{\partial x^p}(0, y) = 0, \quad \text{for } p \text{ odd}, \tag{6.112}$$

$$-\frac{\partial^q u}{\partial y^q}(x, 0) = \frac{\partial^q u}{\partial y^q}(x, 0) = 0, \quad \text{for } q \text{ odd}. \tag{6.113}$$

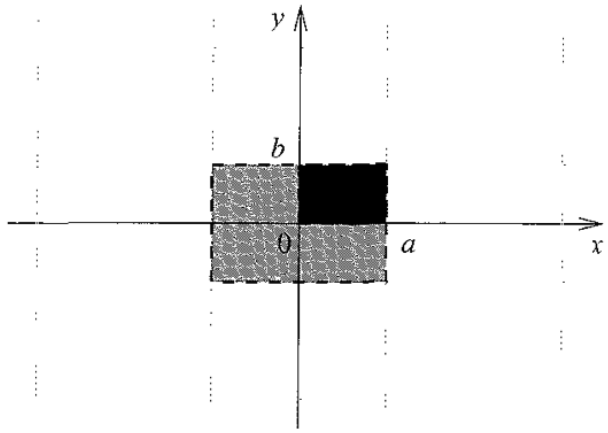


Fig 6 15 The original domain $0 \leq x \leq a, 0 \leq y \leq b$ (shaded black) is extended to the domain shaded grey ($-a \leq x \leq a, -b \leq y \leq b$) by reflection in the boundaries $x = 0$ and $y = 0$. This larger domain is then extended periodically to the lattice with vertices at $(a(1 \pm 2m), b(1 \pm 2n))$, marked by dotted lines.

Using periodicity we also have

$$-\frac{\partial^p u}{\partial x^p}(-a, y) = \frac{\partial^p u}{\partial x^p}(a, y) = \frac{\partial^p u}{\partial x^p}(-a, y) = 0, \quad p \text{ odd}, \tag{6.114}$$

$$-\frac{\partial^q u}{\partial y^q}(x, -b) = \frac{\partial^q u}{\partial y^q}(x, b) = \frac{\partial^q u}{\partial y^q}(x, -b) = 0, \quad q \text{ odd} \tag{6.115}$$

We can now use periodicity to extend the solution to the whole of the plane $\mathbb{R}^2$ by translation of the cell $-a \leq x \leq a, -b \leq y \leq b$ The governing equation is equivariant under reflections and translations, so that

$$f(\gamma u, \mu) = \gamma f(u, \mu) = \gamma \frac{\partial u}{\partial t} = \frac{\partial(\gamma u)}{\partial t}, \tag{6.116}$$

where γ is the reflection or translation, and this means that the new solution is also a solution of the governing equation

This extension procedure embeds the original bifurcation problem within the bifurcation problem on a periodic lattice, $\mathcal{L}$, with vertices at $(a(1 \pm 2m), b(1 \pm 2n))$, for $m = 0, 1, 2, \ldots$, and $n = 0, 1, 2, \ldots$, as shown in Figure 6 15 In order for $\mathcal{L}$ to be a lattice group we have to use translated coordinates $x' = x - a$ and $y' = y - b$ so that $(0, 0) \in \mathcal{L}$: in the new coordinates $\mathcal{L}$ is the lattice generated by the translations $(2a, 0)$ and $(0, 2b)$. From now on we shall use the original coordinates x and y, but when we refer to $\mathcal{L}$ as a group we must bear in mind this implicit shift of the origin. The symmetry group relevant to the bifurcation on the lattice is $D_2 \ltimes T^2$, where D_2 is the symmetry group of the rectangle $-a \leq x \leq a$, $-b \leq y \leq b$, and T^2 is the torus of translations

If we work with solutions that are periodic on the lattice, we can pick out those that satisfy the original reflecting boundary conditions by restricting our attention to $\mathrm{Fix}(\mathcal{D}_2)$, where $\mathcal{D}_2$ is the group generated by the reflections $r_x : x \to -x$ and $r_y : y \to -y$. Solutions that are invariant under these reflections satisfy equations (6.107) and (6.108), and lie in $\mathrm{Fix}(\mathcal{D}_2)$.

Solutions that are periodic on the extended lattice lie in $\mathrm{Fix}(\mathcal{L})$. The simplest such functions are

$$u_{(n_x,n_y)}(x, y) = \cos\left(\frac{n_x x\pi}{a} + \phi_x\right)\cos\left(\frac{n_y y\pi}{b} + \phi_y\right), \tag{6.117}$$

where n_x and n_y are integers and ϕ_x and ϕ_y are real constants. We assume at least one of n_x and n_y is nonzero so that we are considering periodic rather than constant solutions. Any such solution will lie in $\mathrm{Fix}(\mathcal{D}_2)$, and hence satisfy the reflecting boundary conditions on the original domain, if and only if $n_x\phi_x = n_y\phi_y = 0$.

The apparent symmetries of the restricted problem in $\mathrm{Fix}(\mathcal{D}_2)$ are those in $N(\mathcal{D}_2)$, which are generated by the reflections r_x, r_y, m_x and m_y. The hidden symmetries then are the remaining reflections and translations of $\mathcal{D}_2 \ltimes T^2$. The effective symmetry group in the restricted space $\mathrm{Fix}(\mathcal{D}_2)$ is $N(\mathcal{D}_2)/\mathcal{D}_2 \cong \mathcal{D}_2$, with generators m_x and m_y. This is exactly the symmetry group that we naively expected would govern the bifurcation in the original domain.

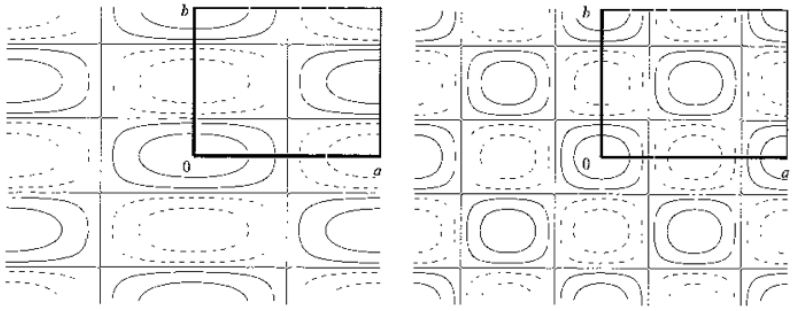
If an experimenter did not want hidden symmetries in his or her system, the reflections at the boundaries could be destroyed by putting a slight bend in the boundaries. This could be done in such a way as to preserve the reflections in the lines $x = a/2$ and $y = b/2$, and hence the D_2 symmetry of the original domain.

6.7.1 Hidden translations

Let us assume that only one mode $u_{(n_x,n_y)}$ is present, in other words that n_x and n_y are fixed up to a change of sign and that this mode is the only one to lie on the critical circle. The translations $x \to x + a/n_x$ and $y \to y + b/n_y$ are elements of the symmetry group, $\mathcal{D}_2 \ltimes T^2$, of the lattice bifurcation problem that are hidden symmetries in $\mathrm{Fix}(\mathcal{D}_2)$ if $n_x > 1$ and $n_y > 1$. Applying either of them in the flow-invariant subspace $\mathrm{Fix}(\mathcal{D}_2)$ we find $u_{(n_x,n_y)} \to -u_{(n_x,n_y)}$. So equivariance under these translations implies that a mode $u(x, y, t) = A(t)\cos(n_x x\pi/a)\cos(n_y y\pi/b)$ will arise at a pitchfork bifurcation where the real amplitude, $A(t)$, evolves according to the equation

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \mu A + cA^3 + \cdots \tag{6.118}$$

with c a real constant. If either n_x or n_y is odd, this result does not come as a surprise, because the symmetry of the original domain $x \to a - x$ or $y \to b - y$



(a) (b)

Fig 6.16. (a) There are no hidden symmetries when the original rectangular domain (outlined by a thick line on the contour plots) contains an odd number of half-wavelengths of the pattern in one or both directions (b) In contrast if a whole number of wavelengths fit into the domain in both directions, there are hidden translation symmetries

respectively leads to $u_{(n_x\ n_y)} \to -u_{(n_x,n_y)}$, and so the requirement for equivariance under these reflections would lead to an equation of the form (6.118). If both n_x and n_y are even, however, then $u_{(n_x\ n_y)}$ is invariant under both the reflection symmetries of the original domain, and so in the first instance one would assume that the equivariance requirements were trivially satisfied, and that the mode would arise at a transcritical bifurcation So in the case where n_x and n_y are both even, the hidden translation symmetries $x \to x + a/n_x$ and $y \to y + b/n_y$ alter the expected form of the amplitude equation

Figure 6 16 shows that the half-wavelength translation $x \to x + a/n_x$ ($y \to y + b/n_y$) has the same effect as the reflection symmetry $x \to a - x$ ($y \to b - y$) if n_x (n_y) is odd, but is not equivalent to either reflection symmetry if both n_x and n_y are even.

6.7.2 Hidden rotational symmetries

If there is more than one mode $u_{(n_x\ n_y)}$ on the critical circle, there may be hidden rotational symmetries too. Let $a = 1$ and $b = \sqrt{3}$, and let the critical wavenumber be $k = 2/\sqrt{3}$, where $k^2 = (n_x/a)^2 + (n_y/b)^2$ Now there are two sets of values (n_x, n_y) that lie on the critical circle $n_x = n_y = 1$, and $n_x = 0$, $n_y = 2$, corresponding to the eigenfunctions

$$u_{(1\,1)} = \cos(\pi x + \phi_{1x}) \cos\left(\frac{\pi y}{\sqrt{3}} + \phi_{1y}\right), \tag{6.119}$$

$$u_{(0\,2)} = \cos\left(\frac{2\pi y}{\sqrt{3}} + \phi_{2y}\right) \tag{6.120}$$

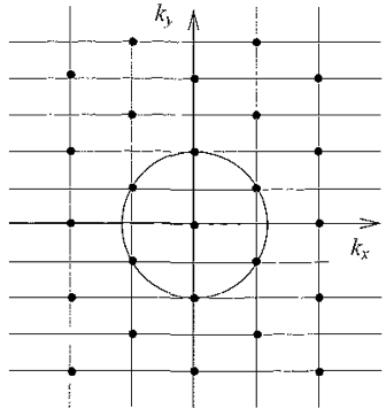


Fig 6.17 Dual lattice showing rotational hidden symmetry The rectangular lattice defined by Neumann boundary conditions on a domain $0 \leq x \leq a, 0 \leq y \leq b$ gives rise to the rectangular dual lattice drawn in solid lines Permitted wavevectors (k_x, k_y) lie at the vertices of the lattice, where $k_x = n_x\pi$, $k_y = n_y\pi/\sqrt{3}$ and n_x and n_y are integers. The black dots pick out a hexagonal dual lattice hidden in the rectangular array

respectively, where ϕ_{1x}, ϕ_{1y} and ϕ_{2y} are phases that can take any real value. These modes correspond to combinations of the six equally spaced modes of the dual lattice lying on the critical circle, as shown in Figure 6.17. The figure shows that there is a hidden hexagonal dual sublattice, corresponding to a hexagonal lattice, $\mathcal{L}'$, with a symmetry of rotation through $2\pi/3$. This hidden symmetry crops up any time we restrict a Euclidean-invariant problem to a rectangular lattice with aspect ratio $1 : \sqrt{3}$ – it is not unique to the case of a rectangular domain with reflecting boundary conditions In fact the possibility of hidden rotations should be borne in mind whenever you restrict a problem to a lattice

Solutions periodic on the rectangular lattice, $\mathcal{L}$, can be written as a combination of the two critical modes as follows:

$$\begin{aligned} u(x, y, t) &= 4A(t)\cos(\pi x + \phi_{1x})\cos\left(\frac{\pi y}{\sqrt{3}} + \phi_{1y}\right) \\ &\quad + 2B(t)\cos\left(\frac{2\pi y}{\sqrt{3}} + \phi_{2y}\right) \qquad (6\,121) \\ &= 2A(t)\cos\left(\pi x + \frac{\pi y}{\sqrt{3}} + \phi_{1x} + \phi_{1y}\right) \\ &\quad + 2A(t)\cos\left(\pi x - \frac{\pi y}{\sqrt{3}} + \phi_{1x} - \phi_{1y}\right) \\ &\quad + 2B(t)\cos\left(\frac{2\pi y}{\sqrt{3}} + \phi_{2y}\right), \qquad (6\,122) \end{aligned}$$

where $A(t)$ and $B(t)$ are real and their scaling is chosen to be convenient in what follows This is a restricted set of the solutions periodic on the hexagonal lattice, $\mathcal{L}'$, that take the form

$$u(x, y, t) = z_1 e^{i\pi(x+y/\sqrt{3})} + z_2 e^{i\pi(-x+y/\sqrt{3})} + z_3 e^{-2\pi i y/\sqrt{3}} + c\ c\ , \qquad (6.123)$$

where $z_i(t) \in \mathbb{C}$, so in this case it is appropriate to take the symmetry group of the bifurcation on the hexagonal lattice, $D_6 \ltimes T^2$, as the large group. The rectangular solutions are given by $z_1 = Ae^{i(\phi_{1x}+\phi_{1y})}$, $z_2 = Ae^{i(\phi_{1x}-\phi_{1y})}$ and $z_3 = Be^{i\phi_{2y}}$

Solutions that satisfy the reflecting boundary conditions on the original domain lie in $\mathrm{Fix}(D_2)$ and are invariant under $x \to -x$ and $y \to -y$, so they must have $\phi_{1x} = \phi_{1y} = \phi_{2y} = 0$, giving

$$u(x, y, t) = 4A \cos \pi x \cos \frac{\pi y}{\sqrt{3}} + 2B \cos \frac{2\pi y}{\sqrt{3}} \qquad (6.124)$$

In $\mathrm{Fix}(D_2)$, the effective symmetries are once more those of the original domain, which act on A and B according to

$$m_x : x \to 1 - x, (A, B) \to (-A, B), \qquad (6.125)$$

$$m_y : y \to \sqrt{3} - y, (A, B) \to (-A, B), \qquad (6.126)$$

and so to cubic order we would expect the following equations for the evolution of A and B:

$$\frac{dA}{dt} = \mu_1 A + \alpha_1 AB + \beta_1 A^3 + \beta_2 B^2 A, \qquad (6.127)$$

$$\frac{dB}{dt} = \mu_2 B + \alpha_2 A^2 + \alpha_3 B^2 + \beta_3 A^2 B + \beta_4 B^3, \qquad (6.128)$$

where the constants μ_i, α_i and β_i are all real

However, the hidden translation $y \to y + \sqrt{3}/2$ acts on the subspace $(0, B)$ according to $(0, B) \to (0, -B)$, and so we must have $\alpha_3 = 0$ The hidden rotation through $2\pi/3$,

$$\rho : u(x, y, t) \to u\left(-\frac{x}{2} + \frac{\sqrt{3}y}{2}, -\frac{\sqrt{3}x}{2} - \frac{y}{2}, t\right), \qquad (6.129)$$

acts trivially on the subspace $A = B$, but since this is $\mathrm{Fix}(\rho)$ it is flow-invariant, and so equations (6.127) and (6.128) must be identical for $A = B$. Consequently we must have $\mu_1 = \mu_2 \equiv \mu$, $\alpha_1 = \alpha_2 \equiv \alpha$ and $\beta_1 + \beta_2 = \beta_3 + \beta_4$.

The resulting amplitude equations are

$$\frac{dA}{dt} = \mu A + \alpha AB + \beta_1 A^3 + \beta_2 B^2 A, \tag{6.130}$$

$$\frac{dB}{dt} = \mu B + \alpha A^2 + (\beta_1 + \beta_2 - \beta_4) A^2 B + \beta_4 B^3. \tag{6.131}$$

These equations should be the restriction to $\mathrm{Fix}(\mathcal{D}_2)$ of the evolution equations for the amplitudes, z_i, on the hexagonal lattice, which take the form we derived in Section 5.4, namely

$$\frac{dz_1}{dt} = \mu z_1 + \alpha \bar{z}_2 \bar{z}_3 - \nu_1 |z_1|^2 z_1 - \nu_2 (|z_2|^2 + |z_3|^2) z_1, \tag{6.132}$$

$$\frac{dz_2}{dt} = \mu z_2 + \alpha \bar{z}_3 \bar{z}_1 - \nu_1 |z_2|^2 z_2 - \nu_2 (|z_3|^2 + |z_1|^2) z_2, \tag{6.133}$$

$$\frac{dz_3}{dt} = \mu z_3 + \alpha \bar{z}_1 \bar{z}_2 - \nu_1 |z_3|^2 z_3 - \nu_2 (|z_1|^2 + |z_2|^2) z_3, \tag{6.134}$$

where α, ν_1 and ν_2 are real constants. The restriction to $\mathrm{Fix}(\mathcal{D}_2)$ is achieved by setting $(z_1, z_2, z_3) = (A, A, B)$, where $A(t)$ and $B(t)$ are real, to get

$$\frac{dA}{dt} = \mu A + \alpha AB - (\nu_1 + \nu_2) A^3 - \nu_2 B^2 A, \tag{6.135}$$

$$\frac{dB}{dt} = \mu B + \alpha A^2 - \nu_1 B^3 - 2\nu_2 A^2 B. \tag{6.136}$$

However equations (6.130) and (6.131) have one more coefficient at cubic order than equations (6.135) and (6.136) so there must be further hidden symmetries restricting the equations in $\mathrm{Fix}(\mathcal{D}_2)$ that we have not used yet.

If we combine a reflection in the line $x = \sqrt{3}y$,

$$x \to \frac{x}{2} + \frac{\sqrt{3}y}{2}, \quad y \to \frac{\sqrt{3}x}{2} - \frac{y}{2}, \tag{6.137}$$

with the translation $y \to y + \sqrt{3}/2$ and call this combined operation ϕ, then using the scalar action of the Euclidean group we transform

$$u(x, y, t) \to u(\phi^{-1}(x, y), t) = u\left(\frac{x}{2} + \frac{\sqrt{3}y}{2} - \frac{3}{4}, \frac{\sqrt{3}x}{2} - \frac{y}{2} + \frac{\sqrt{3}}{4}, t\right), \tag{6.138}$$

which induces the transformation

$$(z_1, z_2, z_3) \to (-iz_1, -z_3, -iz_2). \tag{6.139}$$

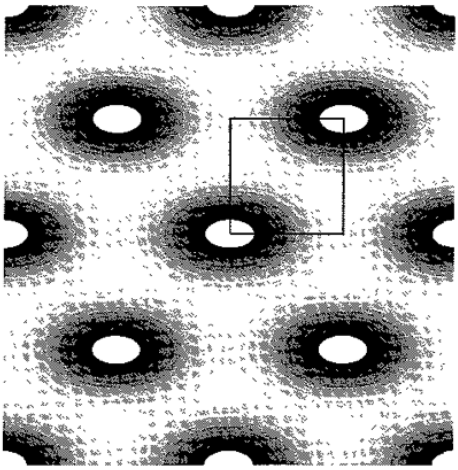


Fig 6 18. Filled contour plot of the solution $u(x, y) = 4\cos\pi x \cos\pi y/\sqrt{3} + 2\cos 2\pi y/\sqrt{3}$. The rectangle marked with solid lines is the original domain $0 \le x \le 1, 0 \le y \le \sqrt{3}$

For $(z_1, z_2, z_3) = (A, A, B) \in \mathrm{Fix}(\mathcal{D}_2)$ we have

$$\phi(z_1, z_2, z_3) = (-iA, -B, -iA), \tag{6 140}$$

and so

$$\mathrm{Fix}(\mathcal{D}_2) \cap \phi\mathrm{Fix}(\mathcal{D}_2) = \{\mathbf{0}\} \tag{6.141}$$

To discover what effect this symmetry has in $\mathrm{Fix}(\mathcal{D}_2)$ we must therefore work in the space, $\mathrm{Fix}(\mathcal{L}')$, of functions periodic on the hexagonal lattice, $\mathcal{L}'$ In this extended space the symmetry ϕ shows that the coefficients of the two terms $|z_i|^2 z_j$ with $i \neq j$ in the evolution equation for z_j must be the same for $i, j = 1, 2, 3$. Combining this with symmetry under rotation through $2\pi/3$ shows that the coefficients must also be the same in all three equations, as in equations (6 132)–(6.134) Restriction to $\mathrm{Fix}(\mathcal{D}_2)$ then shows that the coefficient of A^2B in the $\mathrm{d}B/\mathrm{d}t$ equation must be twice that of B^2A in the $\mathrm{d}A/\mathrm{d}t$ equation This removes the extra degree of freedom in the restricted equations (6.130) and (6 131)

This is another example of a hidden symmetry that does not act on a subspace of the restricted problem, but that none the less affects the evolution there

Recall that solutions with $A = B$ have a hidden symmetry of rotation through $2\pi/3$ You can actually see this in the solutions: for example, if we set $A = B = 1$ then we have the solution

$$u(x, y) = 4\cos\pi x \cos\pi y/\sqrt{3} + 2\cos 2\pi y/\sqrt{3}, \tag{6 142}$$

which is plotted in Figure 6 18. The hexagonal symmetry is clearly visible.

Exercises

6 1 What are the isotropy subgroups and their generators for super hexagons and rolls on the hexagonal superlattice defined by the wavevectors $k_1 = (0, 1)$ and $k_2 = (\sqrt{3}/2, -1/2)$? Are there any hidden symmetries of either of these planforms? What are the other axial planforms on the hexagonal superlattice (in the absence of midplane reflection symmetry)?

6 2 Derive the amplitude equations for a $1:3$ mode interaction problem in a one-dimensional system with $O(2)$ symmetry.

6 3 Derive the amplitude equations relevant to a steady spatial-period-doubling bifurcation.

6 4 Work out the amplitude equations for eightfold quasipatterns, and find the stationary solutions and the phase and phason modes

6 5 Work out the amplitude equations for the steady bifurcation on a square lattice under the pseudoscalar action of $E(2)$. Find the simple anti-square solution and determine its isotropy subgroup.

6 6 Consider the steady bifurcation on the rhombic lattice, using the fundamental representation

$$u(x\ t) = z_1(t)e^{ik_1 x} + z_2(t)e^{ik_2 x} + c.c\ + h\ o\ t$$

Work out the equivalent of Table 5 1 and discuss the role of hidden symmetries

6 7 Show that $N(\mathcal{L}) = H \ltimes \mathbb{R}^2$.

7

Spatial modulation and envelope equations

In Chapter 2 we saw that a system with discrete eigenmodes can be reduced to its evolution on a centre manifold. By requiring patterns to be periodic with respect to a lattice in Chapter 5 we could distinguish between critical and decaying modes and apply the centre manifold theorem to extract amplitude equations for the critical modes. However, if the pattern is not exactly periodic our analysis must allow for the possibility that modes arbitrarily close to the critical modes in Fourier space contribute to the pattern. Then the distinction between stable and unstable modes becomes a little blurred: a mode with growth rate infinitesimally greater than zero will grow, but infinitely slowly, whereas a mode with growth rate infinitesimally less than zero will decay, but again infinitely slowly. In this case, we cannot perform a centre manifold reduction, since we cannot separate the growing and decaying modes well enough. Specifically, we cannot find an appropriate δ in equation (2.30) of Chapter 2. In cases such as these we must use an alternative method of analysis. This chapter describes how envelope equations can be used to describe the evolution of patterns that fit almost, but not exactly, onto a lattice.

7.1 Envelope equations for specific models

As explained in Chapter 5 pattern-forming systems can often be described adequately by a set of partial differential equations for a marker quantity, such as the density or temperature perturbation in a convecting fluid, together with appropriate boundary conditions. In this form, the problem is amenable to analysis using envelope equations. To explain the method, we will look at a specific example.

The Swift–Hohenberg equation (Swift & Hohenberg, 1977) was originally proposed as a simplified model of convective instability in a one-dimensional system.

It takes the form

$$\frac{\partial u}{\partial t} = \left[\mu - \left(\nabla^2 + k_c^2\right)^2\right]u - u^3, \tag{7.1}$$

where $u(x, y, t)$ represents a linear combination of the vertical fluid velocity and the temperature perturbation, μ is a bifurcation parameter, and k_c is a constant. Variations in the vertical direction have been averaged out in this model. We will assume here that the Laplacian includes both horizontal dimensions, and that the system is so large in the two horizontal directions that it is effectively infinite in horizontal extent. The latter point is particularly important, since it means that modes of all wavelengths are permitted, whereas if we were to consider a finite domain, typical boundary conditions such as 'no-slip' ($u = 0$ at $x = 0$ and L and $y = 0$ and L, where L is the length of the box in the two horizontal dimensions), periodic ($u(0, y) = u(L, y)$, $u(x, 0) = u(x, L)$) or 'stress-free' ($\partial u/\partial x$ at $x = 0$ and L and $\partial u/\partial y$ at $y = 0$ and L) would select an infinite set of discrete eigenmodes that fit in the box.

The solution $u = 0$ will become unstable as μ passes through zero. If we consider a mode of the form $u(x, y, t) = \hat{u}e^{\sigma t + i\boldsymbol{k}\cdot\boldsymbol{x}} + c.c.$, where $|\hat{u}| \ll 1$ is a constant and $\boldsymbol{x} = (x, y)$, then linearising equation (7.1) around $u = 0$, gives the following expression for the growth rate σ:

$$\sigma = \mu - \left(k^2 - k_c^2\right)^2, \tag{7.2}$$

where $k^2 = |\boldsymbol{k}|^2$. The most unstable modes, which are the ones with the largest growth rate, have $k^2 = k_c^2$. When μ is negative, all modes have negative growth rate and will decay. At $\mu = 0$ any mode with wavenumber k_c, or any superposition of such modes, regardless of the orientation of the wavevector, can appear and contribute to the pattern since such modes have zero growth rate and will neither grow nor decay. No particular configuration is preferred over any other according to linear theory. Assuming a finite number of modes is present, as is typical for pattern-forming systems, we have

$$u(x, y) = \sum_j A_j e^{i\boldsymbol{k}_j\cdot\boldsymbol{x}} + c.c., \tag{7.3}$$

where $|\boldsymbol{k}_j|^2 = k_c^2$, and the amplitudes A_j are constants. There is no time-dependence in the solution $u(x, y)$ since the contributing modes are stationary at $\mu = 0$. If in fact a continuum of modes were present we would write

$$u(x, y) = \int_{|k|=k_c} \widetilde{A}(k)e^{i\boldsymbol{k}\cdot\boldsymbol{x}}dk + c.c., \tag{7.4}$$

where $\widetilde{A}(k)$ is the Fourier transform of the amplitude, but as we typically want to look at patterns that at onset are made up of a small number of modes we won't pursue the continuum formulation here.

As μ increases, the growth rate σ will become positive for a range of wavenumbers k, and $|u|$ will grow. The observable pattern will be made up of stationary and growing modes with $\sigma \geq 0$. As $|u|$ increases, the nonlinear terms become important, and these eventually select the pattern, stabilising some Fourier modes and causing others to decay.

We will start by analysing the simplest possible pattern, namely rolls or stripes, so we assume here that only one pair of wavevectors arises in the system and we have

$$u(x) = Ae^{ik_c x} + \bar{A}e^{-ik_c x} \tag{7.5}$$

at the bifurcation $\mu = 0$. Both the $\boldsymbol{k} = (+k_c, 0)$ and $\boldsymbol{k} = (-k_c, 0)$ modes are present, since the solution $u(x)$ must be real. We shall see later on that the approach can be extended to more complicated patterns such as squares and hexagons, made up of several superposed Fourier modes.

Close to the bifurcation point, μ is small, and we can write

$$\mu = \epsilon^2 \widehat{\mu}, \tag{7.6}$$

where $|\epsilon| \ll 1$. From equation (7.2), we expect that a band of wavenumbers of width $O(\epsilon)$, centred on $k = k_c$, will now be unstable. For a wavevector $\boldsymbol{k} = (k_c + k_x)\widehat{\boldsymbol{x}} + k_y \widehat{\boldsymbol{y}}$, where $\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{y}}$ are unit vectors in the x and y directions respectively, the growth rate is given by

$$\sigma = \epsilon^2 \widehat{\mu} - \left(2k_c k_x + k_x^2 + k_y^2\right)^2 \tag{7.7}$$

If k_c is $O(1)$, then for all terms in the equation to contribute at the same order in ϵ, consistent scalings are $k_x \sim \epsilon$, $k_y \sim \sqrt{\epsilon}$ and $\sigma \sim \epsilon^2$; this confirms the size of the unstable band, illustrated in Figure 7.1. Since there is a continuous spectrum of permitted wavenumbers – and hence growth rates – right up to the stability boundary, we cannot separate the growing and the decaying modes well enough to perform a centre manifold reduction. We must find an alternative method of analysis to reduce our infinite-dimensional system of modes to something more manageable. In fact we turn to a multiple scales analysis.

The effect of the small band of unstable wavenumbers around $|\boldsymbol{k}| = k_c$, is to modulate the envelope of the carrier wave, $e^{ik_c x}$, so that the amplitude A varies slowly in time and space (Figure 7.2). The appropriate modulation scales can be deduced from the scalings for σ, k_x and k_y: we introduce the slow time variable $T = \epsilon^2 t$ and the slow space variables, $X = \epsilon x$ and $Y = \sqrt{\epsilon} y$, where $0 < \epsilon \ll 1$,

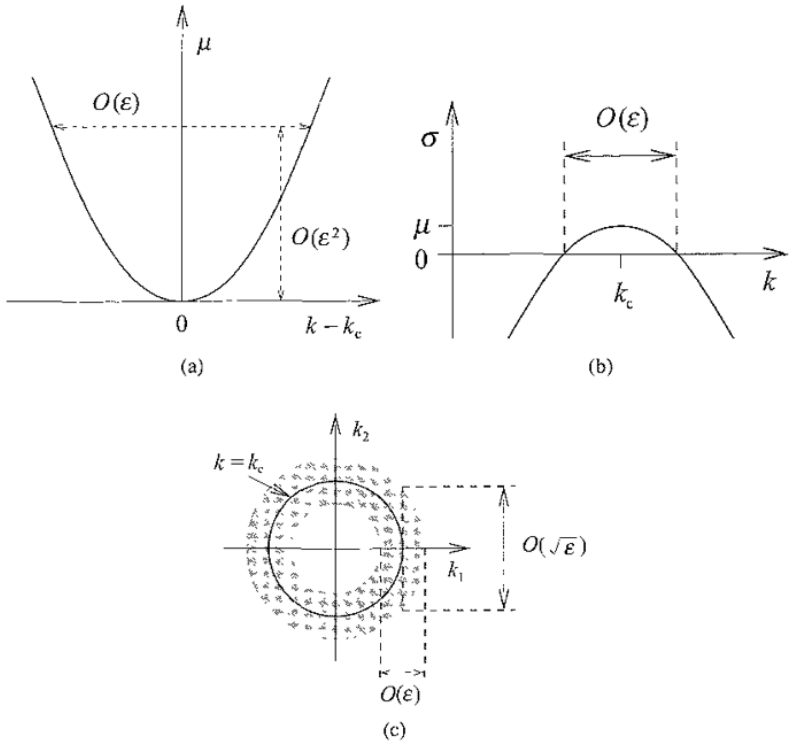


Fig 7 1 (a) The neutral (or marginal) stability curve $\mu = (k^2 - k_c^2)^2$. The curve shows the locus in (μ, k) space of Fourier modes with zero growth rate ($\sigma = 0$) The area above the curve represents growing modes ($\sigma > 0$) that contribute to the observed pattern, and the area below represents decaying modes ($\sigma < 0$) Close to the onset of instability, when the order parameter μ has size $O(\epsilon^2)$, the band of growing modes has width $O(\epsilon)$ (b) The corresponding growth rate, σ, as a function of the wavenumber, k, for a fixed value of μ. Once again the $O(\epsilon)$ band of growing modes with positive growth rates is apparent (c) The growing modes lie in an annulus (shaded grey) in Fourier space, where $\boldsymbol{k} = (k_1, k_2)$. The circle $k = k_c$ is shown as a solid line. The region of growing modes around $\boldsymbol{k} = (k_c, 0)$ has width $O(\epsilon)$ in the k_1 direction and width $O(\sqrt{\epsilon})$ in the k_2 direction

together with the fast scales $\widetilde{t} = t$, $\widetilde{x} = x$ and $\widetilde{y} = y$ We then write

$$u(\widetilde{x}, \widetilde{y}, \widetilde{t}; X, Y, T) = A(X, Y, T)e^{ik_c\widetilde{x}} + c.c., \tag{7.8}$$

and a formal **multiple scales expansion** determines the evolution equation for $A(X, Y, T)$.

To perform the multiple scales analysis, we make the substitutions

$$\mu = \epsilon^2\widehat{\mu}, \quad \frac{\partial}{\partial x} = \frac{\partial}{\partial \widetilde{x}} + \epsilon\frac{\partial}{\partial X}, \quad \frac{\partial}{\partial y} = \frac{\partial}{\partial \widetilde{y}} + \sqrt{\epsilon}\frac{\partial}{\partial Y}, \quad \frac{\partial}{\partial t} = \frac{\partial}{\partial \widetilde{t}} + \epsilon^2\frac{\partial}{\partial T}, \tag{7.9}$$

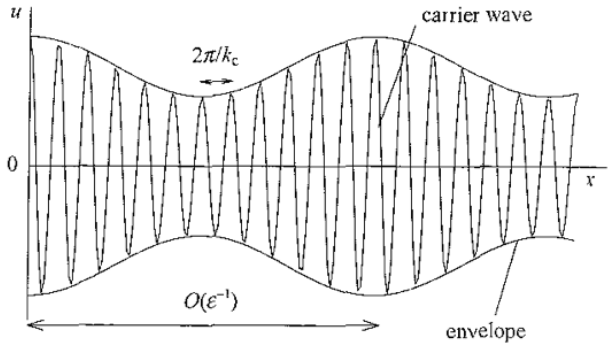


Fig 7.2. The modulated stripe pattern $u(x, y, t) = A(X, Y, T)e^{ik_c x} + \bar{A}(X, Y, T)e^{-ik_c x}$ plotted against x for a fixed value of Y and T. The carrier wave $e^{ik_c x}$ has wavelength $2\pi/k_c$, where k_c is the critical wavenumber for the system. The modulation occurs because a band of wavenumbers of width $O(\epsilon)$ centred around the critical wavenumber k_c also contribute to the pattern, leading the envelope, $A(X, Y, T)$, to vary on a long $O(\epsilon^{-1})$ lengthscale in x.

so that equation (7.1) becomes

$$\epsilon^2 \frac{\partial u}{\partial T} = \epsilon^2 \hat{\mu} u - \left\{ \left(\frac{\partial^2}{\partial \tilde{x}^2} + k_c^2 \right)^2 + 2\epsilon \left(\frac{\partial^2}{\partial \tilde{x}^2} + k_c^2 \right) \left(2\frac{\partial^2}{\partial \tilde{x} \partial X} + \frac{\partial^2}{\partial Y^2} \right) + 2\epsilon^2 \frac{\partial^2}{\partial X^2} \left(\frac{\partial^2}{\partial \tilde{x}^2} + k_c^2 \right) + \epsilon^2 \left(2\frac{\partial^2}{\partial \tilde{x} \partial X} + \frac{\partial^2}{\partial Y^2} \right)^2 + \cdots \right\} u - u^3, \tag{7.10}$$

noting that u has no dependence on $\tilde{y}$ or $\tilde{t}$. At this point we drop the tilde from $\tilde{x}$ since $\tilde{x}$ is just x.

We now expand the solution, u, in powers of ϵ as follows:

$$u = \epsilon u_1 + \epsilon^2 u_2 + \epsilon^3 u_3 + \cdots \tag{7.11}$$

where u_1, u_2 and u_3 are $O(1)$ functions of x, X, Y and T. Often the multiple scales technique will work even if the correct scaling for u can't be guessed in advance: the solution can be expanded in multiples of the lowest power of ϵ appearing in the linear analysis, in this case $\epsilon^{\frac{1}{2}}$, and the appropriate scaling will eventually emerge from the requirement that we end up with a nonlinear evolution equation for A. It will, however, take longer to do the calculations and incorrect scalings will probably lead you up a few blind alleys along the way.

At $O(\epsilon)$, we have

$$L(u_1) \equiv \left(\frac{\partial^2}{\partial x^2} + k_c^2 \right)^2 u_1 = 0 \tag{7.12}$$

If we consider only solutions which are bounded as $x \to \pm\infty$, then we must have

$$u_1(x, y, t) = A(X, Y, T)e^{ik_c x} + c.c\ , \tag{7.13}$$

where we have rescaled $A \to \epsilon A$ compared with equation (7.8). At $O(\epsilon^2)$, we have

$$L(u_2) = -2\left(\frac{\partial^2}{\partial x^2} + k_c^2\right)\left(2\frac{\partial^2}{\partial x \partial X} + \frac{\partial^2}{\partial Y^2}\right)u_1, \tag{7.14}$$

but the righthand side is zero by equation (7.13), so equation (7.14) has the solution $u_2(x, y, t) = B(X, Y, T)e^{ik_c x} + c.c$. We can absorb u_2 into u_1 (in other words $u_1 + \epsilon u_2 \to u_1$) since the new u_1 will still be $O(1)$ at leading order. This is equivalent to setting $B = 0$. The $O(\epsilon^3)$ equation is then

$$L(u_3) = -\frac{\partial u_1}{\partial T} + \hat{\mu}u_1 - 2\frac{\partial^2}{\partial X^2}\left(\frac{\partial^2}{\partial x^2} + k_c^2\right)u_1 - \left(2\frac{\partial^2}{\partial x \partial X} + \frac{\partial^2}{\partial Y^2}\right)^2 u_1 - u_1^3 \tag{7.15}$$

If there are to be no resonant terms in the equation for u_3, so that u_3 is bounded as $x \to \pm\infty$ and the expansion is uniformly valid in time and space, then the coefficients of $e^{\pm ik_c x}$ on the righthand side of equation (7.15) must be zero.

There is another way of looking at this requirement that involves solutions of the adjoint equation to (7.12). The **adjoint operator** $L^\dagger$ is defined by the equation

$$< u|L^\dagger(v) > = < L(u)|v >, \tag{7.16}$$

where $< \ | \ >$ is a suitable scalar product. In this case the scalar product is the integral $\int_0^{2\pi/k_c} dx$, and we have $L^\dagger = L$ under suitable boundary conditions, so L is **self-adjoint**. The **Fredholm alternative** says that either the equation

$$L(u) = 0 \tag{7.17}$$

has no nontrivial solutions or the adjoint homogeneous equation

$$L^\dagger(v) = 0 \tag{7.18}$$

has a nontrivial solution $v \neq 0$. In the second case, the Fredholm alternative theorem states that

$$L(u) = f \tag{7.19}$$

has a solution if and only if

$$< f|v > = 0, \tag{7.20}$$

for all nonzero v satisfying equation (7.18). Clearly if equation (7.19) has a solution then

$$< f|v > = < L(u)|v > = < u|L^{\dagger}(v) > = 0, \tag{7.21}$$

for $v \neq 0$ such that $L^{\dagger}(v) = 0$. The proof that equation (7.19) has a solution if equation (7.20) holds is more involved and involves the use of integral equations (see Porter & Stirling, 1990, for further details). In the current calculation, the operator is self-adjoint ($L^{\dagger} = L$), so $e^{\pm ik_c x}$ are nontrivial solutions of equation (7.18), and we must have

$$\int_0^{2\pi/k_c} e^{\pm ik_c x} f_3(x; X, Y, T)\mathrm{d}x = 0, \tag{7.22}$$

in order for equation (7.15) to have a solution for u_3, where $f_3(x; X, Y, T)$ is defined by the righthand side of equation (7.15)

$$f_3(x; X, Y, T) = -\frac{\partial u_1}{\partial T} + \widehat{\mu} u_1 - 2\frac{\partial^2}{\partial X^2}\left(\frac{\partial^2}{\partial x^2} + k_c^2\right)u_1 - \left(2\frac{\partial^2}{\partial x \partial X} + \frac{\partial^2}{\partial Y^2}\right)^2 u_1 - u_1^3. \tag{7.23}$$

Since f_3 can be expressed as the sum of Fourier modes $e^{ink_c x}$, where n is an integer and since $\int_0^{2\pi/k_c} e^{i(n\pm 1)k_c x}\mathrm{d}x$ is zero for $n \neq \pm 1$ respectively, the only terms that will contribute to the integrals in equation (7.22) are the coefficients of $e^{\pm ik_c x}$. In order for equation (7.22) to hold, both these coefficients must be zero independently. The coefficient of $e^{-ik_c x}$ is in fact the complex conjugate of the coefficient of $e^{ik_c x}$, because equation (7.1) is real, and so there is only one condition to satisfy.

Substituting equation (7.13) into equation (7.15) we see that the coefficient of $e^{ik_c x}$ is

$$-\frac{\partial A}{\partial T} + \widehat{\mu} A - 3|A|^2 A + 4k_c^2\left(\frac{\partial}{\partial X} - \frac{i}{2k_c}\frac{\partial^2}{\partial Y^2}\right)^2 A \tag{7.24}$$

and we must set this equal to zero. In other words, the following **envelope equation** must be obeyed

$$\frac{\partial A}{\partial T} = \widehat{\mu} A - 3|A|^2 A + 4k_c^2\left(\frac{\partial}{\partial X} - \frac{i}{2k_c}\frac{\partial^2}{\partial Y^2}\right)^2 A \tag{7.25}$$

Envelope equations are often referred to simply as amplitude equations, and we shall use both terms interchangeably from now on. The coefficients in the amplitude equation are not free, as we will see that they would be if we had derived the

equation using symmetry methods, but are determined directly from the original governing Swift–Hohenberg equation.

In summary, the velocity, u can be written

$$u(x, y, t) = \epsilon A(X, Y, T)e^{ik_c x} + c.c. + \cdots, \tag{7.26}$$

where the omitted terms are higher harmonics in the Fourier expansion, and where the amplitude, A, of the carrier wave, $e^{ik_c x}$, will evolve according to the amplitude equation (7.25) to leading order. This is a good approximation to the behaviour of the original system close to the onset of instability, as long as the amplitude does not vary rapidly in time or space.

We can set the critical wavenumber to unity, $k_c = 1$, by rescaling the x variable, and the amplitude equation may be renormalised using the transformations, $\widehat{A} = \sqrt{3}A$, $\widehat{X} = X/2k_c$ and $\widehat{Y} = Y/\sqrt{2}$ to give

$$\frac{\partial A}{\partial T} = \mu A - |A|^2 A + \left(\frac{\partial}{\partial X} - \frac{i}{2}\frac{\partial^2}{\partial Y^2}\right)^2 A, \tag{7.27}$$

where all the hats, including the one on μ, have been dropped. This is the **Newell–Whitehead–Segel equation** (Newell & Whitehead, 1969; Segel, 1969), which we will use to investigate the stability of roll patterns in the next chapter. In fact we could also have set $\mu = 1$ by rescaling the time variable T, but it is usual to leave the bifurcation parameter in the equation explicitly for the sake of clarity. For neatness, some authors also remove the factor of $1/2$ in front of the term $-i\partial^2/\partial Y^2$, by scaling Y differently. Retention of the factor of $1/2$ is in some sense 'natural' because it corresponds to scaling x such that $k_c = 1$ and then choosing the scale factor for X to be the square of that for Y; in any case this is the scaling we shall adopt in this book.

Another way of thinking of the dropping of the hat on μ is that we have reabsorbed the scalings with respect to ϵ into the variables, so that we have $\mu \sim \epsilon^2$, $A \sim \epsilon$, $\partial/\partial T \sim \epsilon^2$, $\partial/\partial X \sim \epsilon$ and $\partial/\partial Y \sim \epsilon^{\frac{1}{2}}$. This approach is very common in the literature, and equation (7.26) would often be written without the ϵ in front of the amplitude. In much of the rest of this book the scalings will be implicit in this manner.

7.2 Envelope equations and symmetries

Frequently, a particular pattern occurs in more than one physical context: for example, stripes are seen in convection, in reaction-diffusion systems, and on a zebra's back. In this case, it might be preferable to deduce the amplitude equations directly from the observable features and symmetries of the pattern and its environment,

rather than dealing with one particular physical, chemical or biological system that exhibits the pattern. This symmetry-based approach will be described next. It shows explicitly that the set of allowed patterns depends on the symmetries of the system, and hence that qualitatively similar behaviour may arise in very different systems. Consideration of the symmetries of a system is a very powerful tool for understanding the evolution of patterns that arise in real situations, as we have seen in previous chapters.

In the case of rolls or stripes, we first assume that the pattern can be modelled as a carrier wave with a slowly varying amplitude. Then choosing the x direction to be perpendicular to the roll axis, and scaling x so that the most unstable mode has wavenumber $k_c = 1$ we have

$$u(x, y, t) = A(X, Y, T)e^{ix} + c.c. \tag{7.28}$$

to leading order, where $u(x, y, t)$ measures some physical marker quantity, such as a temperature perturbation in convection or the concentration of a chemical reactant in a reaction-diffusion system, through which the pattern can be observed. The amplitude, A, will turn out to be small, but we make no explicit assumptions about its scaling at this point.

In order to write down the growth rate for a Fourier mode proportional to $e^{\sigma t + ik\cdot x}$, we must consider certain properties of the rolls and their environment. We assume, based on observation of the pattern we want to model that:

(i) the system varies under the influence of an external control parameter μ, for example the Rayleigh number, or amount of heating, in convection;
(ii) the most unstable Fourier mode has $|\boldsymbol{k}|^2 = 1$, and becomes unstable at $\mu = 0$;
(iii) the environment is isotropic, so the growth rate σ does not depend on the direction of the wavevector $\boldsymbol{k}$ and hence must be a function of $k^2 = |\boldsymbol{k}|^2$. (We expect σ to depend on k^2 rather than k since physically reasonable governing equations will give rise to growth rates that are differentiable in Fourier space.) Close to the bifurcation we can therefore expand σ in powers of $(k^2 - 1)$.
(iv) the rolls are steady, so there is no oscillatory behaviour and σ must be real.

With all these requirements in mind, we find that σ must be written

$$\sigma = \mu - \xi_0^2(k^2 - 1)^2 + O([k^2 - 1]^3), \tag{7.29}$$

where ξ_0 is a real constant, that may depend smoothly on μ, and is $O(1)$ at $\mu = 0$. As before the growth rate has a quadratic maximum in $(k^2 - 1)$; there can be no linear term since if there were the maximum would be displaced from $k^2 = 1$.

If we look at Fourier modes close to the critical mode e^{ix}, we can write $\boldsymbol{k} = (1 + k_x, k_y)$ and so the growth rate is given by

$$\sigma = \mu - \xi_0^2\left(2k_x + k_x^2 + k_y^2\right)^2 + O([k^2 - 1]^3). \tag{7.30}$$

Close to the bifurcation at $\mu = 0$, μ is small and we write $\mu = \epsilon^2\widehat{\mu}$, where $0 < \epsilon \ll 1$ and consistent scalings are $k_x = \epsilon\widehat{k}_x$, $k_y = \sqrt{\epsilon}\widehat{k}_y$ and $\sigma = \epsilon^2\widehat{\sigma}$, just as we found for the Swift–Hohenberg model above. Since ξ_0 is $O(1)$ at $\mu = 0$, we can neglect at leading order any dependence it may have on μ. Again we have discovered long space and time scales, suggesting that we may indeed interpret the solution, $u(x, y, t)$, as a carrier wave in a modulated envelope. We introduce slow variables $X = \epsilon x$, $Y = \sqrt{\epsilon}y$ and $T = \epsilon^2 t$, the scalings being given by those for k_x, k_y and σ, and fast variables $\widetilde{x} = x$, $\widetilde{y} = y$ and $\widetilde{t} = t$ as before. The quantity u is written as

$$u(\widetilde{x}, \widetilde{y}, \widetilde{t}; X, Y, T) = A(X, Y, T)e^{i\widetilde{x}} + c.c. + h.o.t., \tag{7.31}$$

and we can immediately drop the tilde from $\widetilde{x}$ as before.

To leading order, equation (7.30) becomes

$$\sigma = \mu - 4\xi_0^2\left(k_x + \frac{1}{2}k_y^2\right)^2 \tag{7.32}$$

We invert the implied Fourier transform by considering a Fourier mode

$$\begin{aligned} u(x; X, Y, T) &= a_0 e^{\widehat{\sigma}\epsilon^2 t + ix + i\widehat{k}_x\epsilon x + i\widehat{k}_y\epsilon^{1/2}y} + c.c., \\ &= a_0 e^{\widehat{\sigma}T + ix + i\widehat{k}_x X + i\widehat{k}_y Y} + c.c., \\ &\equiv A_0(X, Y, T)e^{ix} + c.c., \end{aligned} \tag{7.33}$$

where a_0 is a constant, and $A_0(X, Y, T)$ is a function defined by equation (7.33). Equation (7.32) could clearly have been derived by substituting (7.33) into the linear equation

$$\frac{\partial u}{\partial T} = \widehat{\mu}u + 4\xi_0^2\left(\frac{\partial}{\partial X} - \frac{i}{2}\frac{\partial^2}{\partial Y^2}\right)^2 u, \tag{7.34}$$

and multiplying by the growth rate scaling ϵ^2. Since there is no x dependence in the equation, it also holds for $A_0(X, Y, T)$, and since the amplitude equation should hold for all A, not just the special case $A = A_0$, we conclude that the linear part of the amplitude equation for A must be

$$\frac{\partial A}{\partial T} = \widehat{\mu}A + 4\xi_0^2\left(\frac{\partial}{\partial X} - \frac{i}{2}\frac{\partial^2}{\partial Y^2}\right)^2 A \tag{7.35}$$

The above procedure is equivalent to returning from Fourier space in equation (7.32) to real space via $ik_x \to \partial/\partial X$, $ik_y \to \partial/\partial Y$ and $\sigma \to \partial/\partial T$. (In order to maintain a consistent scaling we must send $\mu \to \hat{\mu}$ at the same time.)

We can also get to this point using a formulation that is continuous in wavenumber space, namely

$$u(x, y, t) = \int \widetilde{A}(k_1, k_2, t) e^{i(k_1 x + k_2 y)} dk_1 dk_2 + c.c. \tag{7.36}$$

The dispersion relation (7.29) implies that to linear order the Fourier transform, $\widetilde{A}(k_1, k_2, t)$, must satisfy

$$\frac{\partial \widetilde{A}}{\partial t} = \mu \widetilde{A} - \xi_0^2 (k^2 - 1)^2 \widetilde{A}. \tag{7.37}$$

Setting $(k_1, k_2) = (1 + \epsilon \widehat{k}_x, \epsilon^{\frac{1}{2}} \widehat{k}_y)$ gives

$$\frac{\partial \widetilde{A}}{\partial T} = \widehat{\mu} \widetilde{A} - \xi_0^2 \left(2\widehat{k}_x + \widehat{k}_y^2\right)^2 \widetilde{A} \tag{7.38}$$

to leading order, where $\mu = \epsilon^2 \widehat{\mu}$ and $T = \epsilon^2 t$ as before, and we also have

$$\begin{aligned} u(x, y, t) &= \int \widetilde{A}(1 + \epsilon \widehat{k}_x, \epsilon^{\frac{1}{2}} \widehat{k}_y, T) e^{ix} e^{i(\epsilon \widehat{k}_x x + \epsilon^{\frac{1}{2}} \widehat{k}_y y)} \epsilon d\widehat{k}_x \, \epsilon^{\frac{1}{2}} d\widehat{k}_y + c.c \\ &= \epsilon^{\frac{3}{2}} e^{ix} \int \widetilde{A}(1 + \epsilon \widehat{k}_x, \epsilon^{\frac{1}{2}} \widehat{k}_y, T) e^{i(\widehat{k}_x X + \widehat{k}_y Y)} d\widehat{k}_x d\widehat{k}_y + c.c \end{aligned} \tag{7.39}$$

We now identify the amplitude or envelope as

$$A(X, Y, T) \propto \int \widetilde{A}(1 + \epsilon \widehat{k}_x, \epsilon^{\frac{1}{2}} \widehat{k}_y, T) e^{i(\widehat{k}_x X + \widehat{k}_y Y)} d\widehat{k}_x d\widehat{k}_y, \tag{7.40}$$

which satisfies the linear envelope equation

$$\frac{\partial A}{\partial T} = \widehat{\mu} A + 4\xi_0^2 \left(\frac{\partial}{\partial X} - \frac{i}{2} \frac{\partial^2}{\partial Y^2} \right)^2 A. \tag{7.41}$$

We do not fix the constant of proportionality in the definition of A at this stage, because the nonlinear terms determine the correct scaling with ϵ.

The nonlinear parts of the amplitude equation are found by considering the symmetries of the system. We assume that the governing equations and boundary conditions have Euclidean symmetry, as does the Swift–Hohenberg equation on an infinite horizontal plane. The zero solution clearly also has Euclidean symmetry. The carrier wave lies on a one-dimensional lattice that inherits some symmetries from the Euclidean group, $E(2)$, namely translation in x and reflection in x. Since we are looking at stripes in a two-dimensional domain we also retain translation

and reflection symmetries in y. (In a one-dimensional domain we simply drop all mention of y and use symmetries in x only to determine the amplitude equation.) If the general form of the solution given in equation (7.31) is to remain valid under these symmetries, the amplitude equation must be equivariant with respect to them.

Translation in x corresponds to

$$x \to x + \phi,\ X \to X + \epsilon\phi,\ A(X, Y, T) \to A(X - \epsilon\phi, Y, T)e^{-i\phi}, \qquad (7.42)$$

where ϕ is a constant. At this point we apply the symmetries $A(X, Y, T) \to A(X, Y, T)e^{-i\phi}$ and $A(X, Y, T) \to A(X - d_x, Y, T)$ separately, where both ϕ and d_x are arbitrary constants. The former, a phase-shift symmetry, is the **normal form symmetry**: in a one-dimensional system, where there is no dependence on Y, it can be shown that the normal form symmetry holds to arbitrarily high order in the expansion of the amplitude equation (Melbourne, 1998). Generically, the two separated parts of the translation symmetry will be broken eventually by terms of the form $A^p\bar{A}^q e^{i(p-q-1)x} \equiv A^p\bar{A}^q e^{i(p-q-1)X/\epsilon}$, for $p + q$ large, and others that involve derivatives of A. The overall translation symmetry $A(X, Y, T) \to A(X - \epsilon\phi, Y, T)e^{-i\phi}$ will however hold to all orders. We usually consider low-order truncations of the amplitude equations, so we do not see the terms that break the normal form symmetry. Since we are working close to onset where $|A|$ is small, the symmetry-breaking terms will be tiny. For most purposes we expect that they will be negligible, but they may become important if we want to use the amplitude equation to examine other small effects, such as the influence of distant boundaries.

A heuristic explanation for splitting the translation symmetry in this way is that we want the amplitude equation to hold for arbitrary small $\epsilon = X/x$. As we vary ϵ, the translation in X, $d_x = \epsilon\phi$, varies for a fixed translation, ϕ, in x. This means that we have to apply the x and X translations independently: invariance under translation in x gives the normal form symmetry, while invariance under translation in X gives the symmetry $A(X, Y, T) \to A(X - d_x, Y, T)$ that mimics the translation symmetry, $u(x, y, t) \to u(x - \phi, y, t)$, of the solution. Of course, in reality the x-dependence of $u(x, y, t)$ cannot be split neatly into exactly two scales corresponding to the variables x and X, and this is reflected in the appearance of terms that break the normal form symmetry at high order.

Translation in y simply gives

$$A(X, Y, T) \to A(X, Y - d_y, T), \qquad (7.43)$$

where d_y is a constant Reflection in x corresponds to

$$X \to -X, \; A \to \bar{A}, \tag{7 44}$$

whereas reflection in y simply corresponds to

$$Y \to -Y. \tag{7 45}$$

The freedom to choose the origin of time is captured by the time-translation symmetry $A(X, Y, T) \to A(X, Y, T - \tau)$ for constant τ Any nonlinear terms which appear in the amplitude equations must be consistent with all the equivariance conditions

Close to the bifurcation point, A will be small, as will its spatial derivatives, so among the possible nonlinear terms, we must look for the one which is of lowest order. The quadratic terms A^2, $|A|^2$, $\bar{A}^2$ are not allowed because they do not transform in the same way as $\partial A/\partial T$ under the normal form symmetry. The same is true of the cubic terms A^3, $|A|^2\bar{A}$ and $\bar{A}^3$. The only cubic term which conforms to this symmetry is $|A|^2A$. This is the lowest order nonlinear term permitted in the amplitude equation; its coefficient is forced to be real by equivariance under reflection in x, and to be independent of X, Y and T by equivariance under space and time translations For details of the enforcement of normal form (or x translation) and reflection symmetry equivariance, look back at Section 5 3, where a very similar procedure is set out step by step.

The lowest order amplitude equation thus turns out to be

$$\frac{\partial A}{\partial T} = \widehat{\mu}A + 4\xi_0^2\left(\frac{\partial}{\partial X} - \frac{i}{2}\frac{\partial^2}{\partial Y^2}\right)^2 A - g|A|^2A, \tag{7 46}$$

where g is a real constant The linear growth rate has a scaling $O(\epsilon^2)$, so for the nonlinear term to come in at the same order, the amplitude must scale as $O(\epsilon)$ and g as $O(1)$ True to form we set $A = \epsilon\widehat{A}$ and immediately drop the hat If the nonlinearity is to **saturate** or counteract the linear instability, then g must also be positive Otherwise, higher order nonlinear terms will have to be introduced until one of them is capable of saturating the instability. Alternatively, the zero solution would be nonlinearly unstable at all orders, which would mean that no ordered pattern would be observed over long times, and we know this is not the case in a system where rolls are observed to form spontaneously. In contrast to the case of the explicit reduction from the Swift–Hohenberg equation, the coefficients in the amplitude equation are not determined in the symmetry-based derivation Depending on your perspective, this is either an advantage, because it makes clear what the general case is, or a disadvantage, because you can't immediately use the equation to describe a particular application.

Again we can renormalise A to set $g = 1$, rescale the space variables to remove the factor of $4\xi_0^2$, and drop the hat from the μ to produce

$$\frac{\partial A}{\partial T} = \mu A - |A|^2 A + \left(\frac{\partial}{\partial X} - \frac{i}{2}\frac{\partial^2}{\partial Y^2}\right)^2 A, \tag{7.47}$$

which is the Newell–Whitehead–Segel equation (7.27) as found above. In the case where we have no Y dependence the equation reduces further to

$$\frac{\partial A}{\partial T} = \mu A - |A|^2 A + \frac{\partial^2 A}{\partial X^2}. \tag{7.48}$$

This is the **Ginzburg–Landau equation** (Ginzburg & Landau, 1950), which is very well known and has applications in many areas of physics and beyond. Ginzburg–Landau theory stems originally from a phenomenological model for superconductivity introduced by Ginzburg and Landau in 1950. (As I write Ginzburg has just been awarded a Nobel prize for his work on the theory of superconductors.) Equation (7.48) is sometimes called the real Ginzburg–Landau equation, because its coefficients are real, to distinguish it from the complex Ginzburg–Landau equation, which has complex coefficients (see Section 7.5). In both cases the amplitude, A, is complex in general.

The linear terms in the amplitude equation must also satisfy the equivariance conditions, and it is easy to check that they do. This is no accident – we required the growth rate given in equation (7.29) to respect Euclidean symmetry by insisting that there should be no dependence on the direction of the wavevector. Then when we expanded around the carrier wave mode to get equation (7.30) we inherited its reflection and translation symmetries. In fact we could have used equivariance to write down the linear terms of the amplitude equation too. If we had done this we would have found that the linear terms should be

$$\begin{aligned}\frac{\partial A}{\partial T} = {} & a_1 A + i a_2 \frac{\partial A}{\partial X} + a_3 \frac{\partial^2 A}{\partial X^2} + a_4 \frac{\partial^2 A}{\partial Y^2} + i a_5 \frac{\partial^3 A}{\partial X^3} + i a_6 \frac{\partial^3 A}{\partial X \partial Y^2} \\ & + a_7 \frac{\partial^4 A}{\partial X^4} + a_8 \frac{\partial^4 A}{\partial X^2 \partial Y^2} + a_9 \frac{\partial^4 A}{\partial Y^4} + \cdots,\end{aligned} \tag{7.49}$$

where reflection symmetry prohibits the appearance of odd Y-derivatives of A, and forces the coefficients a_j to be real. The rotation symmetry forces the derivatives to appear in the combination $(\partial/\partial X - (i/2)\partial^2/\partial Y^2)$, or in other words imposes the scaling $X \sim Y^2$. The requirement for the most unstable mode to be $\boldsymbol{k} = (1, 0)$ then forces $a_2 = a_4 = 0$ and $a_3 > 0$, so the leading order linear terms are exactly those that appear in the Newell–Whitehead–Segel equation (7.47).

Generically, the Ginzburg–Landau equation in one spatial dimension is valid to describe the evolution of the critical mode envelope, A, close to a stationary

bifurcation in a system with Euclidean symmetry, where the growth rate of Fourier modes has a quadratic maximum at some nonzero wavenumber, $k = k_c \neq 0$, and all other modes are damped at onset

Imagine that now we have an anisotropic system, so that there is no symmetry under rotations, but that $\boldsymbol{k} = (1, 0)$ is still the fastest growing mode. Then the leading order growth rate is

$$\sigma = \mu - \xi_x^2 k_x^2 - \xi_y^2 k_y^2, \tag{7.50}$$

where ξ_x and ξ_y are real constants, which translates to

$$\frac{\partial A}{\partial T} = \mu A + \xi_x^2 \frac{\partial^2 A}{\partial X^2} + \xi_y^2 \frac{\partial^2 A}{\partial Y^2} \tag{7.51}$$

The requirement for $\boldsymbol{k} = (1, 0)$ to be fastest growing means that there can be no linear term in k_x in the growth rate, so that forces $a_2 = 0$, but the X and Y derivatives need no longer appear in any particular combination, so a_4 is no longer forced to be zero. By rescaling X and Y we can set $\xi_x^2 = \xi_y^2 = 1$, and if we then add in the nonlinear term we have, to leading order,

$$\frac{\partial A}{\partial T} = \mu A - |A|^2 A + \nabla^2 A, \tag{7.52}$$

the two-dimensional version of the Ginzburg–Landau equation. It might seem somewhat paradoxical that in an anisotropic environment the evolution of rolls is governed by the isotropic equation (7.52), while in an isotropic system rolls obey the anisotropic Newell–Whitehead–Segel equation, but by picking a direction for the carrier wave we break the symmetry in the isotropic case, forcing the along- and across-roll derivatives to scale differently. In the anisotropic case on the other hand we can scale each direction independently and so bring the equation into isotropic form at leading order.

Equations (7.47), (7.48) and (7.52) all admit roll solutions of the form $A = Re^{iqX}$ where $R^2 = \mu - q^2$ and q is a constant. The roll wavenumber differs from the critical wavenumber $k_c = 1$ by an amount ϵq. We shall see in the next chapter how the nonlinear stability of rolls depends on their deviation from the critical preferred wavelength.

There is a caveat to the use of envelope equations that I should point out now. Neither multiple scales analysis nor the use of symmetry provides a rigorous derivation of an envelope equation. The use of the Ginzburg–Landau equation to model the evolution of the amplitude $A(X, T)$ in the solution $u(x, t) = A(X, T)e^{i\lambda} + c.c.$ in a one-dimensional Euclidean-symmetric system with nonzero preferred wavenumber, $k = k_c \neq 0$, can be justified rigorously (see Melbourne, 1998; 1999). However, in two or more dimensions, the rotational

symmetry of a Euclidean-symmetric system means that the choice of direction for the preferred wavevector is arbitrary. Furthermore, the envelope around the arbitrarily selected preferred mode, or carrier wave, contains other wavevectors on the critical circle that would be expected to grow at the same rate as the carrier wave, and so to be equally important. Though a finite number of isolated critical modes tends to be selected at the nonlinear stage, so that we see lattice patterns, their appearance destroys the original Euclidean symmetry of the problem. Thus the Newell–Whitehead–Segel equation, for example, cannot be derived rigorously. What this means in practice is that we cannot be sure that the solutions of higher-dimensional modulation equations approximate the behaviour of the systems they are intended to model. We can look on them as phenomenological models, bearing in mind that they may fail to capture some aspects of the behaviour that we are interested in. On the other hand, the Newell–Whitehead–Segel equation in particular has been used for many years to describe the evolution of convection rolls in high Prandtl number fluids, and successfully reproduces, close to onset, the stability criteria that are found directly from the governing Boussinesq equations (see, for example, Decker & Pesch, 1994). (At low Prandtl numbers, convection is influenced by the presence of large-scale mean drift flows, and the Newell–Whitehead–Segel equation alone no longer captures all the instabilities of rolls – a measure of the mean flow must be included as an additional dependent variable as we shall see in Chapter 8. In this case the assumption that only modes with wavenumber $k_c \neq 0$ arise at the onset of the instability is violated, because the large-scale mean flow is present with $k = 0$. Similar phenomena are possible in other pattern-forming systems.)

7.3 Free energies or Lyapunov functionals

The real Ginzburg–Landau equation has a **free energy**, or **Lyapunov functional**, given by

$$\mathcal{F}(A, \bar{A}) = \int_{X_-}^{X_+} -\mu|A|^2 + \frac{1}{2}|A|^4 + \left|\frac{\partial A}{\partial X}\right|^2 \mathrm{d}X, \tag{7.53}$$

where the domain of the pattern-forming system is $X_- \leq X \leq X_+$, and where we can let $X_- \to -\infty$ and/or $X_+ \to +\infty$ as necessary. The integrand is known as the **free energy density** and in this case it is given by

$$f(A, \bar{A}) = -\mu|A|^2 + \frac{1}{2}|A|^4 + \left|\frac{\partial A}{\partial X}\right|^2 \tag{7.54}$$

Loosely speaking, a **functional** is a function of functions. In this case $\mathcal{F}$ is a function of A and $\bar{A}$, which themselves are functions of X and T. Differentiating

equation (7.53) with respect to T gives

$$\frac{\mathrm{d}\mathcal{F}}{\mathrm{d}T} = -\int_{X_-}^{X_+} (\mu A - |A|^2 A)\frac{\partial \bar{A}}{\partial T} + (\mu \bar{A} - |A|^2 \bar{A})\frac{\partial A}{\partial T}\mathrm{d}X + \int_{X_-}^{X_+} \frac{\partial A}{\partial X}\frac{\partial^2 \bar{A}}{\partial T \partial X} + \frac{\partial \bar{A}}{\partial X}\frac{\partial^2 A}{\partial T \partial X}\mathrm{d}X. \tag{7.55}$$

We now assume that A satisfies boundary conditions such that

$$\left[\frac{\partial A}{\partial X}\frac{\partial \bar{A}}{\partial T}\right]_{X_-}^{X_+} = 0 \tag{7.56}$$

For example, this will hold if the pattern amplitude is zero outside some central region of the domain. Then integrating the second integral in equation (7.55) by parts gives

$$\frac{\mathrm{d}\mathcal{F}}{\mathrm{d}T} = -\int_{X_-}^{X_+} \left(\mu A - |A|^2 A + \frac{\partial^2 A}{\partial X^2}\right)\frac{\partial \bar{A}}{\partial T}\mathrm{d}X - \int_{X_-}^{X_+} \left(\mu \bar{A} - |A|^2 \bar{A} + \frac{\partial^2 \bar{A}}{\partial X^2}\right)\frac{\partial A}{\partial T}\mathrm{d}X, \tag{7.57}$$

$$= -2\int_{X_-}^{X_+} \left|\frac{\partial A}{\partial T}\right|^2 \mathrm{d}X \leq 0. \tag{7.58}$$

The free energy is always decreasing unless $A(X)$ is a stationary solution of equation (7.48), so the real Ginzburg–Landau equation does not admit time-periodic solutions. If we define the average free energy to be

$$< \mathcal{F} >= \frac{\mathcal{F}}{X_+ - X_-}, \tag{7.59}$$

taking the limit as $X_- \to -\infty$ and/or $X_+ \to +\infty$ if necessary, then $< \mathcal{F} >$ is always decreasing, and is bounded below, since

$$\begin{aligned} < \mathcal{F} > &= \frac{1}{X_+ - X_-}\int_{X_-}^{X_+} -\mu|A|^2 + \frac{1}{2}|A|^4 + \left|\frac{\partial A}{\partial X}\right|^2 \mathrm{d}X, \\ &\geq \frac{1}{X_+ - X_-}\int_{X_-}^{X_+} -\mu|A|^2 + \frac{1}{2}|A|^4 \mathrm{d}X \\ &\geq \frac{1}{X_+ - X_-}\int_{X_-}^{X_+} -\frac{\mu^2}{2}\mathrm{d}X = -\frac{\mu^2}{2}, \end{aligned} \tag{7.60}$$

where we have used

$$g(z) \equiv -\mu z + \frac{1}{2}z^2 \geq -\frac{\mu^2}{2} \tag{7.61}$$

Since $< \mathcal{F} >$ is decreasing and bounded below it must tend to a stationary value, $\mathcal{F}_0$, for example, and this corresponds to a stationary solution for A. Thus solutions of the real Ginzburg–Landau equation must be steady at long times.

The solution with the lowest possible free energy, $< \mathcal{F} >= -\mu^2/2$, is $A = \sqrt{\mu} e^{i\theta_0}$, where θ_0 is a constant, and corresponds to rolls at critical wavenumber. This solution must be stable to all small-amplitude perturbations, because any disturbance would raise the free energy of the system.

If a Lyapunov functional exists, the amplitude equation can be written in the **variational** or **gradient** form

$$\frac{\partial A}{\partial T} = -\frac{\delta \mathcal{F}}{\delta \bar{A}}, \tag{7.62}$$

where $\delta \mathcal{F}/\delta \bar{A}$ is a Fréchet, or variational, derivative, defined such that

$$\int_{X_-}^{X_+} \frac{\delta \mathcal{F}}{\delta \bar{A}} \bar{\eta} \, \mathrm{d}X = \lim_{\epsilon' \to 0} \frac{1}{\epsilon'} \left\{ \mathcal{F}(A, \bar{A} + \epsilon' \bar{\eta}) - \mathcal{F}(A, \bar{A}) \right\}, \tag{7.63}$$

for arbitrary $\eta(X, T)$ such that $A + \epsilon' \eta$ satisfies the boundary conditions. Now if $\mathcal{F}(A, \bar{A})$ is as defined in equation (7.53) then we have

$$\int_{X_-}^{X_+} \frac{\delta \mathcal{F}}{\delta \bar{A}} \bar{\eta} \, \mathrm{d}X = \int_{X_-}^{X_+} -\mu A \bar{\eta} + |A|^2 A \bar{\eta} + \frac{\partial A}{\partial X} \frac{\partial \bar{\eta}}{\partial X} \, \mathrm{d}X, \tag{7.64}$$

$$= \int_{X_-}^{X_+} -\left(\mu A - |A|^2 A + \frac{\partial^2 A}{\partial X^2} \right) \bar{\eta} \, \mathrm{d}X, \tag{7.65}$$

$$= -\int_{X_-}^{X_+} \frac{\partial A}{\partial T} \bar{\eta} \, \mathrm{d}X, \tag{7.66}$$

integrating by parts under suitable boundary conditions as before, and using equation (7.48). Since η is arbitrary we must have

$$\frac{\delta \mathcal{F}}{\delta \bar{A}} = -\frac{\partial A}{\partial T}, \tag{7.67}$$

so we have verified equation (7.62) for the real Ginzburg–Landau equation.

With the amplitude equation in variational form it is an immediate deduction that $\mathcal{F}(A, \bar{A})$ decreases according to

$$\frac{\mathrm{d}\mathcal{F}}{\mathrm{d}T} = \int_{\mathcal{D}} \frac{\delta \mathcal{F}}{\delta A} \frac{\partial A}{\partial T} + \frac{\delta \mathcal{F}}{\delta \bar{A}} \frac{\partial \bar{A}}{\partial T} \mathrm{d}a = -2 \int_{\mathcal{D}} \left| \frac{\partial A}{\partial T} \right|^2 \mathrm{d}a, \tag{7.68}$$

where $\mathcal{D}$ is the system domain, and $\mathrm{d}a$ is the line or area element in one or two dimensions respectively.

The Newell–Whitehead–Segel equation also has a free energy, given by

$$\mathcal{F}(A, \bar{A}) = \int_{\mathcal{D}} -\mu|A|^2 + \frac{1}{2}|A|^4 + \left|\left(\frac{\partial}{\partial X} - \frac{i}{2}\frac{\partial^2}{\partial Y^2}\right) A\right|^2 \mathrm{d}X\mathrm{d}Y, \tag{7.69}$$

where $\mathcal{D}$ is the system domain. The average free energy $< \mathcal{F} >$ is decreasing, and is bounded below by $-\mu^2/2$ as before, so at long times the Newell–Whitehead–Segel equation also has steady roll solutions at critical wavenumber, and does not admit time-periodic solutions.

7.4 Conservation of 'angular momentum'

Stationary solutions of the real Ginzburg–Landau equation,

$$\frac{\partial A}{\partial T} = \mu A - |A|^2 A + \frac{\partial^2 A}{\partial X^2}, \tag{7.70}$$

can be characterised in terms of an 'angular momentum'. Writing $A = Re^{i\theta}$, cancelling a factor of $e^{i\theta}$ and separating the real and imaginary parts of the equation leads to

$$\frac{\partial R}{\partial T} = \mu R - R^3 + \frac{\partial^2 R}{\partial X^2} - R\left(\frac{\partial \theta}{\partial X}\right)^2, \tag{7.71}$$

$$R\frac{\partial \theta}{\partial T} = 2\frac{\partial R}{\partial X}\frac{\partial \theta}{\partial X} + R\frac{\partial^2 \theta}{\partial X^2}. \tag{7.72}$$

Setting $\partial/\partial T = 0$, and multiplying the second equation by R gives

$$\frac{\mathrm{d}}{\mathrm{d}X}\left(R^2\frac{\mathrm{d}\theta}{\mathrm{d}X}\right) = 0, \tag{7.73}$$

and so the 'angular momentum'

$$h = R^2\frac{\mathrm{d}\theta}{\mathrm{d}X} \tag{7.74}$$

is constant throughout the domain. This conservation law is a consequence of equivariance of the amplitude equation under translations in x. It means that there can be no smooth changes of wavenumber from one part of the domain to another without accompanying changes in the amplitude. It also indicates that if $R = 0$ at any point in the domain, then at any other point where the amplitude is nonzero, the phase gradient, $\mathrm{d}\theta/\mathrm{d}X$, must be zero. This is perfect wavenumber selection: only one wavenumber is permitted throughout the entire domain. This situation can arise on a **subcritical ramp** in the control parameter, where $\mu(X)$ varies in space from negative to positive. In the region where $\mu(X)$ is negative, only the zero solution is stable, so we have $R = 0$ and hence $h = 0$. The angular momentum h

is conserved in space, so we must have $d\theta/dX = 0$ even in the region $\mu(X) > 0$, where rolls are stable. The critical wavenumber is therefore selected throughout the region of pattern formation. For a more detailed analysis of rolls on a ramp, see Kramer *et al.* (1982) and Pomeau and Zaleski (1983)

We can see more by substituting equation (7.74) into equation (7.71) in the form

$$\frac{d^2R}{dX^2} = -\frac{dV}{dR} \tag{7.75}$$

for stationary solutions, where

$$V(R) = \frac{1}{2}\mu R^2 - \frac{1}{4}R^4 + \frac{h^2}{2R^2}. \tag{7.76}$$

Multiplying equation (7.75) by dR/dX and integrating with respect to X gives

$$\frac{1}{2}\left(\frac{dR}{dX}\right)^2 + V(R) = E, \tag{7.77}$$

where E is a constant

The behaviour of solutions can be captured in phase-space diagrams of dR/dX against R, and depends on the values of μ and h.

Let us consider the case $\mu > 0$ first. Stationary roll solutions of the form $A = R_0 e^{iqX}$, with $R_0^2 = \mu - q^2$, are fixed points in phase space, since $R = R_0$ and $dR/dX = 0$. For such a solution to exist we must have $0 < q^2 < \mu$, and then equation (7.74) gives

$$h(q) = q(\mu - q^2) \tag{7.78}$$

As shown in Figure 7.3, there are two solutions q_1 and q_2 for a given value of h in the range $0 < |h| < h_{\max}$, where

$$h_{\max} = 2\left(\frac{\mu}{3}\right)^{\frac{3}{2}}, \tag{7.79}$$

is the maximum of h in the region $q^2 < \mu$. For $|h| > h_{\max}$ there are no solutions, q, corresponding to stationary roll solutions, and for $|h| = h_{\max}$, there is exactly one solution, q_*. The case $h = 0$ is special and we shall investigate it separately below.

The phase-space diagrams in Figures 7.4a and b show that for $0 < |h| < h_{\max}$ there are space-periodic solutions for R, corresponding to periodic or quasi-periodic solutions for $u = Ae^{ix}$, while for $|h| > h_{\max}$ there are none. The bold contours in Figure 7.4a each include a homoclinic orbit that corresponds to a compression-dilatation wave that we will discuss further in Chapter 8 in the context of the Eckhaus instability. According to equation (7.74) all solutions where

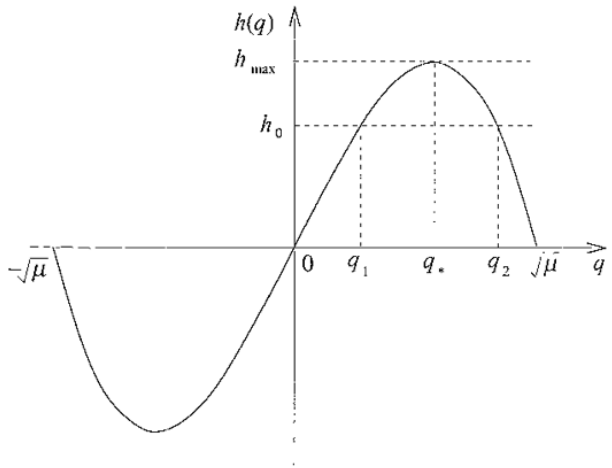


Fig. 7 3 The equation $h(q) \equiv q(\mu - q^2) = h_0$ has two solutions q_1 and q_2 for a given constant value, h_0, in the range $0 < |h_0| < h_{\max}$, where $h_{\max} = 2(\mu/3)^{\frac{3}{2}}$ is the maximum of $h(q)$ in the region $q^2 < \mu$ For $|h| > h_{\max}$ there are no solutions, and for $|h| = h_{\max}$ there is only one solution, q_*

R varies in space gain or lose phase with respect to the carrier wave e^{iX} as X increases The compression-dilatation wave is special in that it tends to the same stationary roll solution $R = R_0 \equiv \sqrt{\mu - q^2}$, as $X \to \pm\infty$ All contours outside the bold contour in Figure 7.4a and all contours in Figure 7 4b correspond to solutions that are unbounded in space, since they have $R \to \pm\infty$ as $X \to \pm\infty$ Orbits with $R < 0$ correspond to patterns phase-shifted by half a spatial period of the carrier wave with respect to those with $R > 0$.

When $\mu < 0$ and $h \neq 0$, the phase-space diagram looks similar to Figure 7 4b, and there are no stationary bounded solutions

In the special case $h = 0$, either R is zero everywhere, and there is no pattern or we have

$$\frac{d\theta}{dX} = 0, \tag{7 80}$$

throughout the region of pattern formation In the latter case, the phase, θ, is constant, and we can set it to zero by choosing the origin of x appropriately So without loss of generality we can choose A to be real, $A = R$, with R satisfying

$$0 = \mu R - R^3 + \frac{d^2R}{dX^2}. \tag{7.81}$$

In the case $\mu > 0$, solutions include the spatially uniform steady states $A = R = \pm\sqrt{\mu}$ and the defect solutions $A = R = \pm\sqrt{\mu}\tanh(\sqrt{\mu/2}X)$ Each defect solution corresponds to a heteroclinic orbit joining the two uniform states in phase

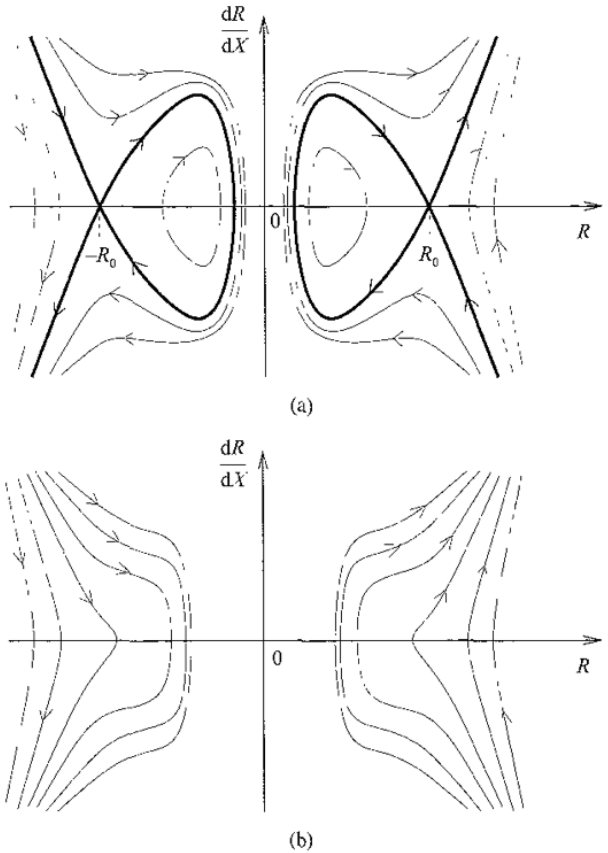
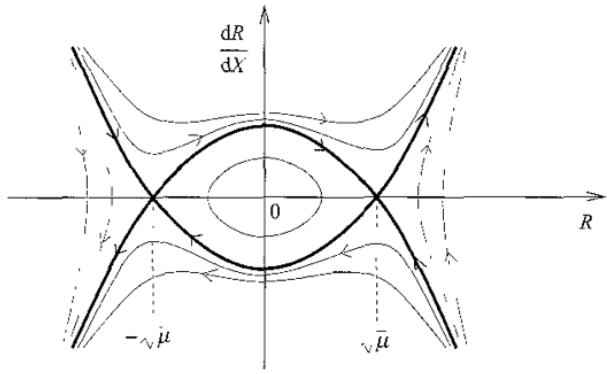


Fig 7 4 Sketched phase space diagrams foı stationary solutions of the Ginzburg–Landau equation (7 70) in the cases (a) $\mu > 0$, $0 < |h| < h_{\max}$, (b) $\mu > 0$, $|h| > h_{\max}$, (c) $\mu > 0$, $h = 0$ and (d) $\mu < 0$, $h = 0$ The bold contour in (a) includes a homoclinic orbit corresponding to a compression/dilatation solitary wave The fixed points at $R = \pm R_0 \equiv \pm\sqrt{\mu - q^2}$ correspond to roll/stripe solutions. The bold contour in (c) includes a heteroclinic orbit connecting the two uniform-state fixed points, $R = \pm\sqrt{\mu}$, and corresponding to a defect solution

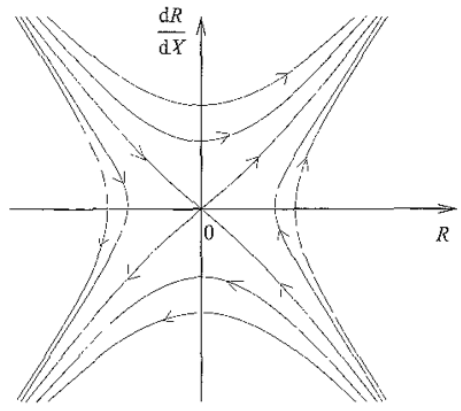
space (see Figure 7.4c). There is a gain or loss of π in the phase between $X \to -\infty$ and $X \to +\infty$. This is equivalent to a defect in the pattern, as shown in Figure 7.5 We can also look for small-amplitude solutions to equation (7.81), linearising around $A = 0$ to get

$$0 = \mu R + \frac{\mathrm{d}^2 R}{\mathrm{d}X^2}, \tag{7.82}$$

which has the solution $R = a\sin(\sqrt{\mu}X) + b\cos(\sqrt{\mu}X)$, where a and b are real constants, corresponding to a spatially periodic or quasiperiodic leading-order solution for $u(x)$

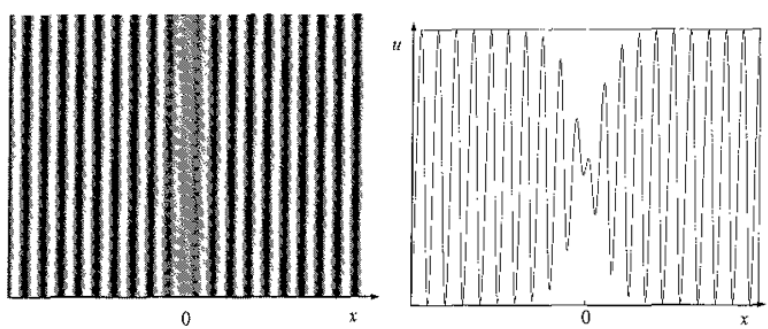


(c)



(d)

Fig 7 4 (*cont.*)



(a) (b)

Fig 7 5 A defect solution $A = \sqrt{\mu}\tanh(\sqrt{\mu/2}X)$ to the real Ginzburg–Landau equation, with $\mu = 1$ The plots show the solution including carrier wave, $u(x) = 2A\cos(x)$, as (a) a two-dimensional greyscale plot, and (b) a graph of $u(x)$ against x, where $X = \epsilon x$ and $\epsilon = 1/2\pi$. The solution gains π in phase as x passes through zero

When $h = 0$ and $\mu < 0$ there are no bounded solutions, as shown in Figure 7.4d.

Note that trajectories can cross the $R = 0$ axis in the case $h = 0$, which cannot happen for $h \neq 0$. The system is symmetric under $R \to -R$, the transformation corresponding to a change of π in the phase, θ, so for $h \neq 0$ we need only consider $R > 0$. The phase space diagrams are shown for both positive and negative values of R to allow comparison with the $h = 0$ case.

7.5 Hopf bifurcations and the complex Ginzburg–Landau equation

So far in this chapter we have focused on spatially modulated roll or stripe patterns that arise at stationary bifurcations from a uniform steady state. Spatially modulated structures can also result from Hopf bifurcations, in which case they will be oscillatory. More complicated spatially modulated patterns, such as squares and hexagons, are also possible. In the remainder of the chapter we will look at some such examples.

In certain systems, such as the Belousov–Zhabotinsky chemical reaction (see Chapter 1), a spatially homogeneous oscillation arises at a Hopf bifurcation from the trivial solution. This oscillation can be modulated on long scales just as in the case of rolls, and we can then write the marker quantity $u(x, y, t)$ in the form

$$u(x, y, t) = A(X, Y, T)e^{-it} + c.c. + h.o.t., \tag{7.83}$$

where X, Y and T are slow scales as before, and where we assume that time has been scaled so that the critical frequency of oscillation, ω_c, is equal to unity. In the case of a chemical reaction u might be a measure of a chemical concentration. We will assume that the system is isotropic and homogeneous, so we have symmetry under reflections and rotations and under translations of the origin in space and time. Under these conditions, the growth rate for a Fourier mode $e^{\sigma t + i\boldsymbol{k}\cdot\boldsymbol{x}}$ with $|\boldsymbol{k}| \ll 1$ and $\boldsymbol{x} = (x, y)$, must take the form

$$\sigma = \mu - \xi_0^2 k^2 + i(-1 + \omega_0 + \omega_1 k^2) + O(k^4), \tag{7.84}$$

where μ is a real bifurcation parameter and where ξ_0, ω_0 and ω_1 are real constants that may depend on μ. The expansion in terms of $k^2 = |\boldsymbol{k}|^2$ is forced by the isotropy, and the form of the expansion ensures that the most unstable mode has critical wavenumber zero ($k = 0$). We must have $\omega_0 = 0$ at $\mu = 0$ so that the critical frequency is unity ($\sigma = -i$ at the onset $\mu = 0$). The remaining parameters ξ_0 and ω_1 are assumed to be $O(1)$ at $\mu = 0$, so to leading order we can neglect any μ-dependence they might have.

The linear parts of the amplitude equation are found by inverting the Fourier transform as for rolls and rescaling, giving

$$\frac{\partial A}{\partial T} = (\mu + i\omega_0)A + (1 + i\alpha)\nabla^2 A, \tag{7.85}$$

where α is a real constant. Here we have used the inversion $(\sigma + i) \to \partial/\partial I$ for the long time scale I, where the $+i$ in the bracket takes account of the basic oscillation e^{-it} in the solution (7.83).

The nonlinear terms must satisfy the requirement that the equation is equivariant under the symmetries of the system. In this case the reflection symmetry $(x, y) \to (-x, y)$, rotation symmetry $\boldsymbol{x} \to R_\theta \boldsymbol{x}$ for some rotation matrix R_θ, and translation in space $\boldsymbol{x} \to \boldsymbol{x} + \boldsymbol{\phi}$ (where $\boldsymbol{\phi}$ is a constant vector) transform $\boldsymbol{X} = (X, Y)$ in the corresponding ways, but imply no further transformation of A. From these symmetries we simply find that the coefficients of the amplitude equation should be independent of $\boldsymbol{X}$. We must also satisfy the time translation symmetry $t \to t + \tau$, $I \to I + \epsilon^2\tau$, where τ is a constant, $I = \epsilon^2 t$ and $|\epsilon| \ll 1$. This symmetry is split into a phase-shift part, corresponding to translation in t, that gives $A \to Ae^{i\tau}$, and a symmetry $A(X, Y, I) \to (X, Y, I - \epsilon^2\tau)$ that corresponds to translations in I. This splitting is analogous to that used for the x translations in the real Ginzburg–Landau equation. The lowest-order term satisfying the phase-shift symmetry is $|A|^2 A$. The coefficient must be independent of $\boldsymbol{X}$ and I to satisfy space and time translation equivariance. We choose its real part to be -1, by rescaling A as necessary, so that the instability is saturated by the nonlinear term. Hence, to leading order, the amplitude equation is

$$\frac{\partial A}{\partial T} = (\mu + i\omega_0)A - (1 + i\beta)|A|^2 A + (1 + i\alpha)\nabla^2 A, \tag{7.86}$$

where α and β are real constants. The term $i\omega_0 A$ can be removed by the transformation $A \to Ae^{i\omega_0 T}$. This is equivalent to changing the frequency of the basic oscillation to $1 - \epsilon^2\omega_0$. The amplitude therefore evolves according to the complex Ginzburg–Landau equation

$$\frac{\partial A}{\partial T} = \mu A - (1 + i\beta)|A|^2 A + (1 + i\alpha)\nabla^2 A, \tag{7.87}$$

to leading order. Consistent scalings close to the onset of instability are given by μ, $\partial/\partial T \sim \epsilon^2$ and $\partial/\partial X$, $\partial/\partial Y$, $A \sim \epsilon$, with α and β both $O(1)$.

In contrast to the real Ginzburg–Landau and Newell–Whitehead–Segel equations, the complex Ginzburg–Landau equation has no Lyapunov functional. This means that time-periodic solutions are not forbidden. Given that we have derived the equation in order to model Hopf bifurcations in a spatially extended system, this should come as no surprise! Among the solutions of the equation are travelling

plane waves given by

$$A = ae^{i\boldsymbol{q}\cdot\boldsymbol{X}-i\omega T}, \tag{7.88}$$

$$a^2 = \mu - q^2 > 0, \tag{7.89}$$

$$\omega = \mu\beta + (\alpha - \beta)q^2, \tag{7.90}$$

where $\boldsymbol{q}$ is a constant vector and $q^2 = |\boldsymbol{q}|^2$. These plane waves crop up again in Chapter 10 when we consider spiral and target patterns.

7.6 Travelling waves and the nonlinear Schrödinger equation

Travelling and standing waves are commonly found in natural systems, for example in rivers and the sea, in beams of light and in chains of sand dunes moving through the desert. We can model waves in one spatial dimension with a slowly varying envelope by considering the quantity

$$u(x,t) = A(X,T)e^{i(t-x)} + B(X,T)e^{i(t+x)} + c.c. + h.o.t., \tag{7.91}$$

where X and T are slow scales and where x and t have been scaled so that the wave has critical wavenumber $k_c = 1$ and critical frequency $\omega_c = 1$. We include both left- and right-travelling waves in order to exploit all the symmetries of the system, as will be seen below. For simplicity we will not consider the two-dimensional case where A and B could depend on an additional long-scale variable Y, though the calculation can be extended to account for this possibility if desired.

Assuming a homogeneous, isotropic environment, the growth rate of a Fourier mode $e^{\sigma t+ikx}$ at close to critical wavenumber, $|k^2 - 1| \ll 1$, is given by

$$\sigma = \mu - \xi_0^2(1-k^2)^2 + i\{1 + \omega_0 + \omega_1(1-k^2) + \omega_2(1-k^2)^2\} + O([1-k^2]^3), \tag{7.92}$$

where μ is the real bifurcation parameter describing the external forcing associated with the instability mechanism, and ξ_0 and the ω_j for $j = 0, 1, 2$, are real constants that can depend on μ. As before, we have $\omega_0 = 0$ at $\mu = 0$ so that $\omega_c = 1$, while the remaining parameters ξ_0, ω_1 and ω_2 are assumed to be $O(1)$ at $\mu = 0$ so their μ-dependence, if any, can be neglected to leading order.

Equation (7.92) is a **dispersion relation**: it relates the growth rate of a Fourier mode to its wavenumber. The imaginary part of the growth rate gives the frequency of oscillation

$$\omega(k) = Im(\sigma(k)) = 1 + \omega_0 + \omega_1(1-k^2) + \omega_2(1-k^2)^2 + \cdots \tag{7.93}$$

In more traditional applications only equation (7.93) would be referred to as the dispersion relation – the real part of σ would be considered separately. The

frequency $\omega(k)$ is related to the speed of propagation of the mode, since if $\omega(k) \equiv kc_p(k)$ is real then $e^{ik(x-c_p t)}$ is a wave travelling to the right with **phase velocity** c_p. Often $c_p(k)$ is not a constant, and so different Fourier modes travel at different speeds. If a wave is made up of different Fourier components, and they travel at different speeds, the original wave will **disperse** – in other words, it will break up. This is where the name 'dispersion relation' comes from. Physical waves will often disperse as a result of some property of the material that they are travelling through, which is then referred to as a **dispersive medium**. The phase velocity is the speed of propagation of a fixed phase of the wave, for example the peak or the trough, and is given by $c_p(k) = \omega(k)/k$. The **group velocity**, defined as $c_g(k) = \mathrm{d}\omega(k)/\mathrm{d}k$, measures the speed of propagation of a wavepacket made up of a range of different wavenumbers, in other words the speed of propagation of the envelope of the wave. To see that the propagation speed of the envelope is $\mathrm{d}\omega(k)/\mathrm{d}k$, consider a wave made up of two Fourier modes with wavenumbers and frequencies that are very close together,

$$u(x,t) = A_0(e^{i(kx-\omega t)} + e^{i\{(k+\delta k)x-(\omega+\delta\omega)t\}} + \text{c.c.}), \tag{7.94}$$

$$= A_0(\cos(kx-\omega t) + \cos\{(k+\delta k)x - (\omega+\delta\omega)t\}), \tag{7.95}$$

$$= 4A_0 \cos\left\{\left(k+\frac{\delta k}{2}\right)x - \left(\omega+\frac{\delta\omega}{2}\right)t\right\} \cos\left(\frac{\delta k}{2}x - \frac{\delta\omega}{2}t\right), \tag{7.96}$$

where $|\delta k| \ll 1$ and $|\delta\omega| \ll 1$ and where A_0 is a real constant. The wave packet is made up of a carrier wave $\cos\left\{\left(k+\frac{\delta k}{2}\right)x - \left(\omega+\frac{\delta\omega}{2}\right)t\right\}$ and an envelope $4A_0\cos\left(\frac{\delta k}{2}x - \frac{\delta\omega}{2}t\right)$. The speed of the envelope, the group velocity, is $\delta\omega/\delta k \approx \mathrm{d}\omega/\mathrm{d}k$. If there is no dispersion, so that $c_p(k)$ is a constant, then the phase and group velocities are equal, but this is not normally the case. If you are not familiar with the basic concepts of wave propagation and would like some further background information you might find the book by Billingham and King (2000) helpful.

For a right-travelling wave $e^{(i+\eta)t-i(1-q)x}$, where $|q| \ll 1$, the dispersion relation (7.92) gives

$$\eta = \mu - 4\xi_0^2 q^2 + i\{\omega_0 + 2\omega_1 q - (\omega_1 - 4\omega_2)q^2\} + O(q^3) \tag{7.97}$$

Inverting the Fourier transform gives the linear part of the amplitude equation for A:

$$\frac{\partial A}{\partial T} = (\mu + i\omega_0)A + 2\omega_1\frac{\partial A}{\partial X} + \{4\xi_0^2 + i(\omega_1 - 4\omega_2)\}\frac{\partial^2 A}{\partial X^2}. \tag{7.98}$$

Following a similar procedure for the left-travelling wave $e^{(i+\eta)t+i(1+q)x}$ we recover the linear part of the amplitude equation for B:

$$\frac{\partial B}{\partial T}=(\mu+i\omega_0)B-2\omega_1\frac{\partial B}{\partial X}+\{4\xi_0^2+i(\omega_1-4\omega_2)\}\frac{\partial^2 B}{\partial X^2}. \tag{7.99}$$

Rescaling X and renaming the coefficients, we can rewrite the linear parts of the equations in the form

$$\frac{\partial A}{\partial T}=(\mu+i\omega_0)A-c\frac{\partial A}{\partial X}+(1+i\alpha)\frac{\partial^2 A}{\partial X^2}, \tag{7.100}$$

$$\frac{\partial B}{\partial T}=(\mu+i\omega_0)B+c\frac{\partial B}{\partial X}+(1+i\alpha)\frac{\partial^2 B}{\partial X^2}, \tag{7.101}$$

where μ is a real bifurcation parameter and ω_0, c and α are real constants

The nonlinear terms are found from symmetry requirements: the set of amplitude equations must be equivariant under

(i) reflection in x: $X\to -X$, $A\leftrightarrow B$;
(ii) translation in x: $x\to x+\phi$, $A\to Ae^{i\phi}$, $B\to Be^{-i\phi}$, ϕ constant;
(iii) translation in X: $X\to X+\Phi$, $A(X,T)\to A(X-\Phi,T)$, $B(X,T)\to B(X-\Phi,T)$, Φ constant;
(iv) translation in t: $t\to t+\psi$, $A\to Ae^{-i\psi}$, $B\to Be^{-i\psi}$, ψ constant;
(v) translation in T: $T\to T+\Psi$, $A(X,T)\to A(X,T-\Psi)$, $B(X,T)\to B(X,T-\Psi)$, Ψ constant.

The lowest-order nonlinear terms that satisfy these conditions are $|A|^2A$ and $|B|^2A$ in the $\partial A/\partial T$ equation and $|A|^2B$ and $|B|^2B$ in the $\partial B/\partial T$ equation. The coefficients must be independent of X and T owing to space and time translation equivariance requirements. The coefficients of $|A|^2A$ and $|B|^2B$ must be the same owing to reflection equivariance and we set the real parts to -1 to saturate the linear instability. Likewise reflection equivariance forces the coefficients of $|B|^2A$ and $|A|^2B$ to be the same. Finally, the form of the nonlinear terms allows the removal of the $i\omega_0A$ and $i\omega_0B$ linear terms using the transformation $A\to Ae^{i\omega_0T}$, $B\to Be^{i\omega_0T}$, which is equivalent to changing the frequency of the carrier wave to $1+\epsilon^2\omega_0$. The full evolution equations to leading order are then a pair of coupled complex Ginzburg–Landau equations,

$$\frac{\partial A}{\partial T}=\mu A-c\frac{\partial A}{\partial X}+(1+i\alpha)\frac{\partial^2 A}{\partial X^2}-(1+i\beta)|A|^2A-(\gamma_r+i\gamma_i)|B|^2A, \tag{7.102}$$

$$\frac{\partial B}{\partial T}=\mu B+c\frac{\partial B}{\partial X}+(1+i\alpha)\frac{\partial^2 B}{\partial X^2}-(1+i\beta)|B|^2B-(\gamma_r+i\gamma_i)|A|^2B, \tag{7.103}$$

where μ is a real bifurcation parameter and c, α, β, γ_r and γ_i are real constants

These equations are in fact somewhat unsatisfactory, because there is no consistent scaling that retains all the terms for $O(1)$ values of the group velocity, c. If we scale μ, $\partial/\partial T \sim \epsilon^2$ and $\partial/\partial X$, A, B, $\sim \epsilon$ as usual, then for consistency we must have $c \sim \epsilon$ and α, β, γ_r and γ_i must be $O(1)$. Strictly, therefore, this analysis is only valid for small group velocities. For amplitude equations derived from governing partial differential equations, the requirement that c be small will usually correspond to a restriction on the spatial wavenumber and temporal frequency of the carrier wave, typically forcing the wavenumber, k_c, to be small. If, on the other hand, c is permitted to be $O(1)$ then we can rescale so that $\partial/\partial X \sim \epsilon^2$, but then the terms $(1+i\alpha)\partial^2 A/\partial X^2$ and $(1+i\alpha)\partial^2 B/\partial X^2$, which describe dissipation and dispersion, do not appear at leading order. Neglecting these effects entirely is undesirable, as it is unlikely to be realistic for applications, so we could modify the equations, retaining the small second-derivative terms and including all the nonlinear terms up to the same order, $O(\epsilon^5)$. In any case the effects of dissipation and dispersion would still be unrealistically small. Each scaling can be relevant in the right circumstances, but there is no one scaling that captures all the effects. This is a serious shortcoming. The problem can be avoided if we consider waves travelling in one direction only: if we restrict attention to a single right-travelling wave, by setting $B = 0$, and transform to a frame moving with the group velocity, c, (writing $\tilde{X} = X - cT$ and immediately dropping the tilde) then the amplitude, A, evolves according to the equation

$$\frac{\partial A}{\partial T} = \mu A + (1+i\alpha)\frac{\partial^2 A}{\partial X^2} - (1+i\beta)|A|^2 A \tag{7.104}$$

The transformation to the moving frame removes any difficulty with scaling the equation consistently, since now if μ, $\partial/\partial T \sim \epsilon^2$, then we can scale $\partial/\partial X$, $A \sim \epsilon$, with α and β both $O(1)$.

Equations (7.102) and (7.103) have right-travelling wave solutions $A = Re^{i\omega T - iqX}$, $B = 0$ with $R^2 = \mu - q^2$ and $\omega = -qc - \alpha q^2 - \beta R^2$, where q is a real constant. The solution for left-travelling waves is similar. Standing waves are given by $A = Re^{i\omega T - iqX}$, $B = Re^{i\omega T + iqX}$ with $R^2 = (\mu - q^2)/(1+\gamma_r)$ and $\omega = -qc - \alpha q^2 - (\beta + \gamma_i)R^2$.

In the special case of nonlinear waves that propagate in dispersive media, but do not grow or decay owing to external forcing (the μA and μB terms in equations (7.102) and (7.103)) or to dissipation (the real part of the terms in $\partial^2 A/\partial X^2$ and $\partial^2 B/\partial X^2$), the growth rate of Fourier modes can have no real part, and equation (7.92) takes the form

$$\sigma = i\{1 + \omega_1(1-k^2) + \omega_2(1-k^2)^2\} + O([1-k^2]^3). \tag{7.105}$$

We also then have time-reversal symmetry ($t \to -t$, $T \to -T$, $A \to \bar{B}$, $B \to \bar{A}$) since the waves neither grow nor decay and can propagate in either direction. Applying these changes to the amplitude equations gives

$$\frac{\partial A}{\partial T} = -c\frac{\partial A}{\partial X} + i\alpha\frac{\partial^2 A}{\partial X^2} - i\beta|A|^2A - i\gamma_i|B|^2A, \tag{7.106}$$

$$\frac{\partial B}{\partial T} = c\frac{\partial B}{\partial X} + i\alpha\frac{\partial^2 B}{\partial X^2} - i\beta|B|^2B - i\gamma_i|A|^2B. \tag{7.107}$$

For a right-travelling wave ($B = 0$), in a frame moving with the group velocity, c, the evolution equation for A reduces to the **nonlinear Schrödinger equation**

$$\frac{\partial A}{\partial T} = i\alpha\frac{\partial^2 A}{\partial X^2} - i\beta|A|^2A, \tag{7.108}$$

a variant of the famous Schrödinger equation for the evolution of wavefunctions in quantum mechanics (Schrödinger, 1926a,b,c,d). More generally, the nonlinear Schrödinger equation describes the propagation of nonlinear waves in dispersive media, for example light propagation in optical fibres. It can be used to analyse the Benjamin–Feir instability of travelling waves, as we shall see in Chapter 8.

7.7 Modulated hexagons

The symmetry methods developed above are easily applied to more complicated patterns in two or more spatial dimensions. In fact they are of particular benefit in this situation, eliminating the need to perform a complicated multiple scales analysis. We shall take modulated hexagons as an example, since hexagons are the most frequently observed experimental pattern after rolls.

We can consider a modulated steady hexagonal pattern to consist of three modulated roll patterns superposed at angles of $2\pi/3$ to one another. Choosing the x direction appropriately and scaling the space variables so that the critical wavenumber is given by $k_c = 1$, the measured physical quantity $u(x, y, t)$ can be written in the form

$$\begin{aligned} u(x, y, t) = {} & A(X, Y, T)e^{ix} + B(X, Y, T)e^{i(-x+\sqrt{3}y)/2} \\ & + C(X, Y, T)e^{-i(x+\sqrt{3}y)/2} + c.c. + h.o.t. \end{aligned} \tag{7.109}$$

The envelopes A, B and C are assumed to vary on long modulation scales X, Y and T as for rolls above. It is useful to adopt the notation $X_1 = X$, $X_2 = (-X + \sqrt{3}Y)/2$ and $X_3 = -(X + \sqrt{3}Y)/2$, and to denote the directions perpendicular to the X_i the Y_i in the obvious manner. We now regard A, B and C as functions of X_i, Y_i and T, for $i = 1, 2, 3$ respectively.

Once again we assume that the pattern-forming system has Euclidean symmetry. The linear parts of the equations for $\partial A/\partial T$, $\partial B/\partial T$ and $\partial C/\partial T$ can be recovered from the dispersion relation (7 29) which we rewrite here for convenience:

$$\sigma = \mu - \xi_0^2(k^2-1)^2 + O([k^2-1]^3) \tag{7.110}$$

Considering the Fourier modes $e^{\sigma t + i(1+k_j)x_j + il_j y_j}$ for $j = 1, 2, 3$, where the x_j and the y_j are the short scales corresponding to the X_j and the Y_j and where k_j and l_j are real constants, we see that the growth rate σ in each case is given by

$$\sigma = \mu - \xi_0^2\left(2k_j + k_j^2 + l_j^2\right)^2 + O([k^2-1]^3) \tag{7.111}$$

Close to the bifurcation at $\mu = 0$, we must find a consistent scaling for all three dispersion relations simultaneously. Writing $\mu = \epsilon^2\hat{\mu}$, we see that we must scale $k_j \sim \epsilon, l_j \sim \epsilon$ and $\sigma \sim \epsilon^2$. Interpreting the Fourier modes as carrier waves in modulated envelopes, and introducing slow variables $X_j = \epsilon x_j$, $Y_j = \epsilon y_j$ and $T = \epsilon^2 t$ we then deduce the linear parts of the amplitude equations to be

$$\frac{\partial A}{\partial T} = \mu A + 4\xi_0^2 \frac{\partial^2 A}{\partial X_1^2}, \tag{7.112}$$

$$\frac{\partial B}{\partial T} = \mu B + 4\xi_0^2 \frac{\partial^2 B}{\partial X_2^2}, \tag{7.113}$$

$$\frac{\partial C}{\partial T} = \mu C + 4\xi_0^2 \frac{\partial^2 C}{\partial X_3^2} \tag{7.114}$$

to leading order, where we have immediately dropped the hat from μ.

The permitted nonlinear terms are found using symmetry arguments. As usual, there is symmetry under translations in x and y. Having chosen to represent the solution in hexagonal form, we also inherit the symmetries of the hexagon from the reflection and rotation symmetries of the system. Consequently the set of amplitude equations must be equivariant under the following symmetry operations:

(i) translation in x: $\{x \to x + \phi, A \to Ae^{-i\phi}, B \to Be^{i\phi/2}, C \to Ce^{i\phi/2}\}$, for constant ϕ;
(ii) translation in y: $\{y \to y + \psi, B \to Be^{-i\sqrt{3}\psi/2}, C \to Ce^{i\sqrt{3}\psi/2}\}$, for constant ψ;
(iii) translation in X: $\{X \to X + \Phi, A(X, Y, T) \to A(X - \Phi, Y, T), \ B(X, Y, T) \to B(X - \Phi, Y, T), C(X, Y, T) \to C(X - \Phi, Y, T)\}$, for constant Φ;
(iv) translation in Y: $\{Y \to Y + \Psi, A(X, Y, T) \to A(X, Y - \Psi, T), \ B(X, Y, T) \to B(X, Y - \Psi, T), C(X, Y, T) \to C(X, Y - \Psi, T)\}$, for constant Ψ;
(v) translation in T: $\{T \to T + \tau, A(X, Y, T) \to A(X, Y, T - \tau), \ B(X, Y, T) \to B(X, Y, T - \tau), C(X, Y, T) \to C(X, Y, T - \tau)\}$, for constant τ;
(vi) reflection in x: $\{X \to -X, A \to \bar{A}, B \to \bar{C}, C \to \bar{B}\}$;

(vii) reflection in y: $\{Y \to -Y,\ B \leftrightarrow C\}$;
(viii) rotation through $2\pi/3$: $\{X_1 \to X_2 \to X_3 \to X_1,\ A \to C \to B \to A\}$

We have once more applied the normal form symmetries, in this case (i) and (ii), separately from the X and Y translation symmetries (iii) and (iv).

The leading order nonlinear terms conforming to these symmetries are

$$\bar{B}\bar{C},\ |A|^2A,\ |B|^2A,\ |C|^2A, i\left(\bar{C}\frac{\partial \bar{B}}{\partial X_2} + \bar{B}\frac{\partial \bar{C}}{\partial X_3}\right),\ i\left(\bar{C}\frac{\partial \bar{B}}{\partial X_3} + \bar{B}\frac{\partial \bar{C}}{\partial X_2}\right) \tag{7.115}$$

in the equation for $\partial A/\partial T$ and the equivalent terms in the remaining two equations, deduced using rotational symmetry. The symmetries also force the coefficients of all these terms to be real and independent of time and space. The amplitude equations then take the form

$$\begin{aligned}\frac{\partial A}{\partial T} &= \mu A + a\bar{B}\bar{C} - b|A|^2A - c(|B|^2 + |C|^2)A \\ &\quad + ig\left(\bar{C}\frac{\partial \bar{B}}{\partial X_2} + \bar{B}\frac{\partial \bar{C}}{\partial X_3}\right) + ih\left(\bar{C}\frac{\partial \bar{B}}{\partial X_3} + \bar{B}\frac{\partial \bar{C}}{\partial X_2}\right) + \frac{\partial^2 A}{\partial X_1^2},\end{aligned} \tag{7.116}$$

$$\begin{aligned}\frac{\partial B}{\partial T} &= \mu B + a\bar{C}\bar{A} - b|B|^2B - c(|C|^2 + |A|^2)B \\ &\quad + ig\left(\bar{A}\frac{\partial \bar{C}}{\partial X_3} + \bar{C}\frac{\partial \bar{A}}{\partial X_1}\right) + ih\left(\bar{A}\frac{\partial \bar{C}}{\partial X_1} + \bar{C}\frac{\partial \bar{A}}{\partial X_3}\right) + \frac{\partial^2 B}{\partial X_2^2},\end{aligned} \tag{7.117}$$

$$\begin{aligned}\frac{\partial C}{\partial T} &= \mu C + a\bar{A}\bar{B} - b|C|^2C - c(|A|^2 + |B|^2)C \\ &\quad + ig\left(\bar{B}\frac{\partial \bar{A}}{\partial X_1} + \bar{A}\frac{\partial \bar{B}}{\partial X_2}\right) + ih\left(\bar{B}\frac{\partial \bar{A}}{\partial X_2} + \bar{A}\frac{\partial \bar{B}}{\partial X_1}\right) + \frac{\partial^2 C}{\partial X_3^2},\end{aligned} \tag{7.118}$$

where μ, a, b, c, g and h are real and where the coefficients of the two cross cubic terms are forced to be identical by reflection symmetry just as they were in the unmodulated case discussed in Section 5.4. In order that all the nonlinear terms appear at the same order in ϵ as the linear terms in the equations, we must scale $\{A, B, C, a\} \sim \epsilon$, with $\{b, c, g, h\}$ all $O(1)$. This means that the coefficient, a, of the quadratic terms $\bar{B}\bar{C}$, $\bar{C}\bar{A}$ and $\bar{A}\bar{B}$ must be small, $O(\epsilon)$, if there is to be a consistent scaling that includes nonlinear terms capable of quenching the linear instability, which only appear at cubic order or higher.

If we further introduce the symmetry $u \to -u$ into the system, the coefficients a, g and h are forced to be zero, as in the analogous case of the steady bifurcation on a hexagonal lattice with $u \to -u$ symmetry, discussed in Chapter 5. In Rayleigh–Bénard convection, this symmetry corresponds to reflection in the horizontal midplane, and is relevant to the case of Boussinesq fluids with identical top and bottom boundary conditions. If the symmetry is broken slightly – in this case if the fluid is slightly non-Boussinesq or the top and bottom boundary conditions are slightly different, for example because the two plates have slightly different thermal conductivities – then we would expect a to be small, but nonzero, which is exactly what is required for the leading order quadratic terms to appear at the same order in the amplitude equations as the cubic terms. We would normally then expect the coefficients, g and h, of the remaining quadratic terms also to be small, since they would typically arise from the same quadratic terms in the original governing partial differential equations. If g and h are $O(\epsilon)$ then we can neglect the corresponding terms in the amplitude equations (7.116)–(7.118) to leading order. However, it can happen that a vanishes exactly by virtue of some special property of the governing equations, while g and h remain nonzero, and possibly $O(1)$, in which case the spatial derivative quadratic terms should be retained (see, for example, the discussions of hexagons in Rayleigh–Bénard convection in Kuznetsov, Nepomnyashchy & Pismen, 1995, and Echebarria & Pérez-García, 2001, for further details). There may also be circumstances in which g and h remain $O(1)$ even though a is small (but nonzero), and then the full equations (7.116)–(7.118) should be used. As expected, the parts of the equations that do not involve spatial modulation are just the same as equations (5.46)–(5.48) derived for the steady bifurcation on a hexagonal lattice.

There are various possible stationary solutions to the amplitude equations (7.116)–(7.118), including x-rolls

$$A = Re^{iqX_1}, \quad B = C = 0,$$
$$R^2 = \frac{1}{b}(\mu - q^2), \tag{7.119}$$

and up-hexagons

$$A = Re^{iqX_1}, \quad B = Re^{iqX_2}, \quad C = Re^{iqX_3},$$
$$0 = (b + 2c)R^2 - \left(a - gq + \frac{1}{2}hq\right)R - (\mu - q^2) = 0, \tag{7.120}$$

where in both cases R and q are real constants. The trivial solution and rectangles are also solutions. If q is zero the rolls or hexagons will be at the critical wavelength, whereas if it is nonzero the patterns will be at a wavelength slightly longer

or shorter than critical. Instabilities of off-critical hexagons will be studied using equations (7.116)–(7.118) in Chapter 9.

We have considered the case where hexagons can be described by a set of amplitude equations truncated at $O(\epsilon^3)$. The inclusion of higher-order nonlinear terms would allow additional steady states, including off-critical versions of the patchwork quilt and triangle solutions described in Chapter 5.

This chapter has focused on deriving amplitude equations that describe the evolution of modulated patterns. In the following chapters we shall use these equations to investigate the modulational behaviour of several common patterns including rolls and spiral waves.

Exercises

7.1 Consider modulated stripes

$$u(x,t) = A(X,T)e^{ix} + c.c. + h.o.t.,$$

where X and T are appropriate long scales for x and t. If the system has translation symmetry, but not reflection symmetry, and if the dispersion relation for the growth rate σ of a Fourier mode proportional to $e^{\sigma t+ikx}$ is given by

$$\sigma = \mu - (k^2-1)^4 + O[(k^2-1)^6], \tag{E7.1}$$

what is the leading order envelope equation for the amplitude, A? (Note that the dispersion relation (E7.1) is reflection-symmetric, but that the reflection symmetry is broken at nonlinear order.)

7.2 Work out the envelope equations for modulated squares

$$u(x,y,t) = A(X,Y,T)e^{ix} + B(X,Y,T)e^{iy} + c.c. + h.o.t.,$$

where X, Y and T are appropriate long scales for x, y and t. Assume that the dispersion relation (7.29) holds, and that the system is isotropic and homogeneous.

7.3 Consider stationary solutions of the envelope equations derived in the previous question that take the form $A = Re^{iqX}$, $B = Se^{ipY}$, where R, S, q and p are real constants. Work out expressions for R and S in terms of p and q and describe what the various possible solutions represent.

7.4 Consider a one-dimensional translation-invariant system with no reflection symmetry, where the dispersion relation for a Fourier mode proportional to $e^{\sigma t+ikx}$ is given by

$$\sigma = \mu + ak^2 - bk^4 + ik(c + dk^2),$$

where $a > 0$, $b > 0$, c and d are real constants, and where μ is a real bifurcation parameter. Write the leading order solution $u(x,t)$ in terms of appropriate envelopes and carrier waves, and derive the evolution equations for the envelopes. (This example is relevant to the wind-driven propagation of sand ripples in the desert, for example, where the existence of a prevailing wind direction breaks reflection symmetry.)

8

Instabilities of stripes and travelling plane waves

Patterns that can be modulated on long scales may undergo new instabilities associated with the modulation or with changes in the pattern wavelength. These typically affect the phase of the pattern, and set in when the pattern wavenumber is not optimal. We shall start by studying the so-called universal instabilities of stripes – the Eckhaus, zigzag and cross-roll instabilities – using the envelope equations derived in Chapter 7, and then move on to the more complicated situation where convection rolls at low Prandtl numbers are coupled to a mean flow. At the end of the chapter we discuss the Benjamin–Feir phase instability of travelling plane waves, which is related to the Eckhaus instability of stripes.

8.1 Universal instabilities of stripes

We will consider rolls or stripes, writing the solution in the form

$$u(x, y, t) = A(X, Y, T)e^{ix} + c.c. + h.o.t., \tag{8.1}$$

where the envelope A evolves according to the Newell–Whitehead–Segel equation

$$\frac{\partial A}{\partial T} = \mu A - |A|^2 A + \left(\frac{\partial}{\partial X} - \frac{i}{2}\frac{\partial^2}{\partial Y^2}\right)^2 A \tag{8.2}$$

Except where stated otherwise we assume that the evolution takes place on the infinite horizontal plane, $-\infty < x < \infty$, $-\infty < y < \infty$ so that we can neglect any possible effect of lateral boundary conditions on the pattern.

In order to understand how a non-optimal pattern wavelength can lead to instability, we will consider a perfect roll pattern at slightly off-critical wavenumber, given by

$$A = R_0 e^{iqX}, \tag{8.3}$$

where R_0 and q are constants. Substituting this into equation (8.2) gives the relationship

$$R_0^2 = \mu - q^2. \tag{8.4}$$

This corresponds to a solution

$$u = R_0 e^{i(1+\epsilon q)x}, \tag{8.5}$$

since $X = \epsilon x$, and so the roll wavenumber deviates from critical by an amount ϵq.

Rolls can only exist for $R_0^2 > 0$, so there is a band of permitted wavenumbers around $q = 0$ (corresponding to a roll wavenumber $k = 1$). The marginal, or neutral, stability curve $\mu = q^2$ ($R_0^2 = 0$) at which the rolls bifurcate from the zero solution is identical at leading order to the neutral stability curve in (μ, k) space shown in Figure 7.1 if we set $k_c = 1$, $k = 1 + \epsilon q$ and scale out a factor of $4\epsilon^2$.

To examine the pattern stability, we now add small perturbations to the roll amplitude and phase, writing $A = R_0(1 + r)e^{i(qX+\phi)}$ where $|r|, |\phi| \ll 1$. Substituting this form for the envelope into equation (8.2), linearising in r and ϕ, and taking the real and imaginary parts of the equation gives

$$\frac{\partial r}{\partial T} = -2R_0^2 r - 2q\frac{\partial \phi}{\partial X} + \frac{\partial^2 r}{\partial X^2} + \frac{\partial^3 \phi}{\partial X \partial Y^2} + q\frac{\partial^2 r}{\partial Y^2} - \frac{1}{4}\frac{\partial^4 r}{\partial Y^4}, \tag{8.6}$$

$$\frac{\partial \phi}{\partial T} = 2q\frac{\partial r}{\partial X} + \frac{\partial^2 \phi}{\partial X^2} - \frac{\partial^3 r}{\partial X \partial Y^2} + q\frac{\partial^2 \phi}{\partial Y^2} - \frac{1}{4}\frac{\partial^4 \phi}{\partial Y^4}. \tag{8.7}$$

Looking for long-wave effects, where the spatial derivatives of the perturbation variables are small in comparison with the variables themselves, and taking modes $r = \hat{r}e^{\sigma T + ikX + ilY}$, $\phi = \hat{\phi}e^{\sigma T + ikX + ilY}$ with $|l| \ll 1$, $k \sim l^2$, and $\hat{r}$ and $\hat{\phi}$ real constants, we find

$$\sigma\hat{r} = -2R_0^2\hat{r} - 2iqk\hat{\phi} - k^2\hat{r} - ikl^2\hat{\phi} - ql^2\hat{r} - \frac{l^4}{4}\hat{r}, \tag{8.8}$$

$$\sigma\hat{\phi} = 2iqk\hat{r} - k^2\hat{\phi} + ikl^2\hat{r} + -ql^2\hat{\phi} - \frac{l^4}{4}\hat{\phi}. \tag{8.9}$$

The relative scalings of k and l follow from those of X and Y. Eliminating the arbitrary constants $\hat{r}$ and $\hat{\phi}$ gives a quadratic equation for the growth rate, σ, namely

$$\left(\sigma + 2R_0^2 + k^2 + ql^2 + \frac{l^4}{4}\right)\left(\sigma + k^2 + ql^2 + \frac{l^4}{4}\right) = k^2(2q + l^2)^2. \tag{8.10}$$

There are two roots

$$\sigma_1 = -2R_0^2 + O(k), \tag{8.11}$$

$$\sigma_2 = -k^2\left(1 - \frac{2q^2}{R_0^2}\right) - ql^2 - \frac{l^4}{4} + O(k^3). \tag{8.12}$$

The first growth rate eigenvalue, $\sigma_1 \approx -2R_0^2 < 0$, corresponds to the eigenvector $(\hat{r}, \hat{\phi}) = (1, O(k))$, and so describes the relatively rapid relaxation, on an $O(1)$ timescale, of the amplitude perturbation, r, to its equilibrium value. Rolls are thus stable to amplitude modes. At leading order this is the same result as we would find for rolls on a lattice, since none of the wavenumbers q, k or l appear in the eigenvalue at leading order. The stability derives from the fact that the nonlinear term quenches, or acts against, the growth of the roll amplitude and so fixes it for given values of μ and q.

In contrast, the second eigenvalue, σ_2, corresponds to a truly modulational mode, since it contains q, k and l at leading order (and these cannot appear if the pattern is restricted to a lattice). The eigenvector is

$$\begin{pmatrix} O(k) \\ 1 \end{pmatrix}, \tag{8.13}$$

and so σ_2 describes the slow evolution, on an $O(l^{-2}) = O(k^{-1})$ timescale, of the phase perturbation, ϕ. If there is no Y-dependence, so that $l = 0$, the evolution will be on an $O(k^{-2})$ timescale. A phase instability will occur when $\sigma_2 > 0$. Two different phase instabilities are possible: the Eckhaus, and zigzag instabilities. We shall consider them both in detail shortly.

An alternative method of determining the behaviour of the phase mode is by slaving the rapidly decaying amplitude perturbation to the slowly varying phase. This is the method of adiabatic elimination that we introduced in Chapter 2, and is the so-called **phase approximation** of Haken (1978) and Pomeau and Manneville (1979). In the long wavelength limit, where spatial derivatives of the perturbation variables are very small in comparison with the variables themselves, it is clear from equations (8.6) and (8.7) that the amplitude varies much more rapidly in time than the phase. We can consider the evolution of the pattern as occurring in two distinct stages. In the first stage, the amplitude evolves rapidly towards an equilibrium state, the rate of change only slowing down when the system is very close to the final steady solution. In the second stage, the phase changes very slowly, and the amplitude adjusts rapidly to these changes; the net evolution though is on a slow timescale, since here the phase is driving the motion, with the amplitude following adiabatically. Pomeau and Manneville (1979) showed that an equation for

the evolution of the phase in this slow régime can be found in terms of the phase gradients.

The phase instabilities take place on a slow timescale. In order to capture their behaviour, and discard the rapid amplitude mode, we rescale

$$\frac{\partial}{\partial X} \to \delta \frac{\partial}{\partial X}, \quad \frac{\partial}{\partial Y} \to \delta^{\frac{1}{2}} \frac{\partial}{\partial Y}, \quad \frac{\partial}{\partial T} \to \delta \frac{\partial}{\partial T}, \tag{8.14}$$

where $|\delta| \ll 1$. From equation (8.6) we have

$$\left(2R_0^2 + \delta \frac{\partial}{\partial T} - \delta q \frac{\partial^2}{\partial Y^2} + O(\delta^2)\right) r = \left(-2\delta q \frac{\partial}{\partial X} + O(\delta^2)\right) \phi. \tag{8.15}$$

Operating on both sides with

$$\left(2R_0^2 - \delta \frac{\partial}{\partial T} + \delta q \frac{\partial^2}{\partial Y^2}\right) \tag{8.16}$$

gives

$$r = -\frac{\delta q}{R_0^2} \frac{\partial \phi}{\partial X} + O(\delta^2). \tag{8.17}$$

Substituting this into the rescaled version of equation (8.7) results in the following **phase-diffusion equation**

$$\frac{\partial \phi}{\partial T} = q \frac{\partial^2 \phi}{\partial Y^2} + \delta \left(1 - \frac{2q^2}{R_0^2}\right) \frac{\partial^2 \phi}{\partial X^2} + -\frac{\delta}{4} \frac{\partial^4 \phi}{\partial Y^4} + O(\delta^2) \tag{8.18}$$

Now considering a mode of the form $\phi = \widehat{\phi} e^{\sigma T + ikX + ilY}$, the growth rate of phase disturbances is found to be

$$\sigma_2 = -ql^2 - \delta k^2 \left(1 - \frac{2q^2}{R_0^2}\right) - \delta \frac{l^4}{4} + O(\delta^2), \tag{8.19}$$

which is the same as σ_2 in equation (8.12) above, when the rescaling is taken into account. In the absence of Y-dependence ($l = 0$) the appropriate time scaling is $\partial/\partial T \sim \delta^2$, which would give

$$\sigma_2 = -k^2 \left(1 - \frac{2q^2}{R_0^2}\right) + O(\delta^3), \tag{8.20}$$

as expected.

For finding the phase behaviour, slaving is often preferable to a full analysis of both rapidly and slowly varying modes, since it is much simpler to carry out in practice; it concentrates solely on the slowly varying phase modes, and all the unnecessary information is discarded early on.

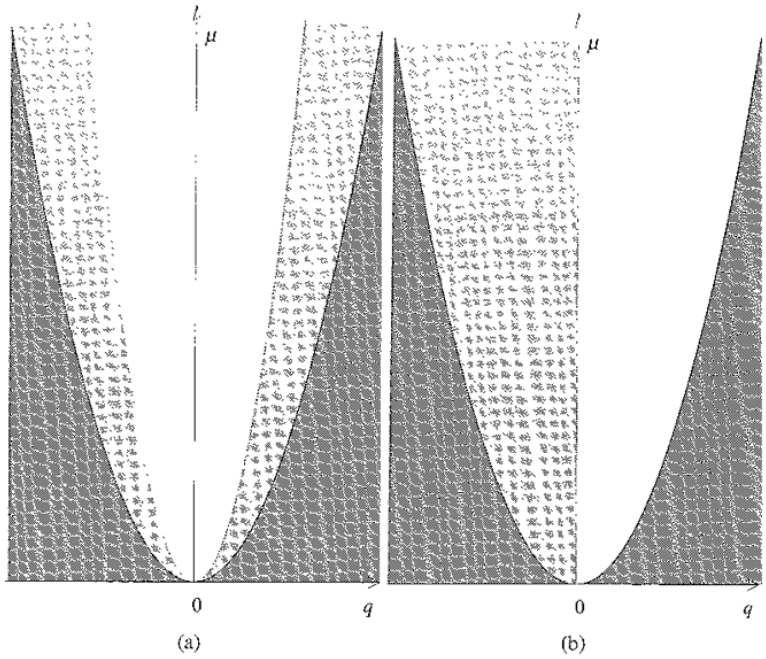


Fig 8.1. The Eckhaus and zigzag instability boundaries for rolls governed by the Newell–Whitehead–Segel equation (8.2). (a) The Eckhaus unstable region for rolls is shaded light grey. It is bounded below by the marginal stability curve, $\mu = q^2$, and above by the Eckhaus stability boundary, $\mu = 3q^2$. (b) The zigzag unstable region for rolls is shaded light grey. It is bounded below by the marginal stability curve, $\mu = q^2$, and to the right by the zigzag stability boundary. In both cases, rolls do not exist in the dark grey region below the marginal stability curve, while in the unshaded regions rolls exist and are stable to (a) the Eckhaus instability or (b) the zigzag instability.

8.2 The Eckhaus instability

For perturbations that vary in the X (across-axis) direction only, we have $l = 0$ in equation (8.12) and the phase perturbation growth rate is

$$\sigma_2 = -k^2\left(1 - \frac{2q^2}{R_0^2}\right) + O(k^4), \tag{8.21}$$

which is positive for rolls in the region $R_0^2 < 2q^2$ ($\mu < 3q^2$) shown in Figure 8.1a. This is the Eckhaus instability (Eckhaus 1965; Newell & Whitehead 1969; Kramer & Zimmermann 1985) which acts on the roll phase to change the wavelength, compressing or dilating the pattern. It occurs when the roll wavelength is too long or too short, and leads eventually to the elimination or creation of rolls at defects in the pattern, and thereby an adjustment to a more favourable wavelength.

A weakly nonlinear analysis close to the onset of the Eckhaus instability leads to an equation for the diffusion of the phase. First we redefine ϕ for ease in the calculations, by writing

$$A = Re^{iq(X+\phi)}, \tag{8.22}$$

where R is the amplitude of the envelope, A, and where the phase perturbation, ϕ, has been rescaled by q. Now we set $q^2 = R_0^2/2 + d\delta^2$, where $|\delta| \ll 1$ and d measures the distance from onset, which corresponds to $\mu = 3R_0^2/2 + d\delta^2$, and then rescale so that $\partial/\partial X \to \delta\partial/\partial X$, $\partial/\partial T \to \delta^4\partial/\partial T$ and $\phi \to \delta\phi$. We are considering patterns that vary with X only, so we also set $\partial/\partial Y = 0$. Then from the real and imaginary parts of equation (8.2) we obtain

$$\delta^4\frac{\partial R}{\partial T} = R_0^2R - R^3 - (R_0^2/2 + d\delta^2)\left(2\delta^2\frac{\partial\phi}{\partial X} + \delta^4\left(\frac{\partial\phi}{\partial X}\right)^2\right)R + \delta^2\frac{\partial^2 R}{\partial X^2}, \tag{8.23}$$

$$\delta^4 R\frac{\partial\phi}{\partial T} = \delta^2 R\frac{\partial^2\phi}{\partial X^2} + 2\frac{\partial R}{\partial X}\left(1 + \delta^2\frac{\partial\phi}{\partial X}\right). \tag{8.24}$$

We expand the amplitude and phase variables in a power series

$$R = R_0 + \delta^2 R_2 + \delta^4 R_4 + \cdots, \tag{8.25}$$

$$\phi = \phi_0 + \delta^2\phi_2 + \delta^4\phi_4 + \cdots, \tag{8.26}$$

and substitute these expressions into equations (8.23) and (8.24), equating terms at each order in δ^2 successively. At $O(1)$, both equations are automatically satisfied. Next at $O(\delta^2)$, both equations are satisfied if

$$R_2 = -\frac{R_0}{2}\frac{\partial\phi_0}{\partial X} \tag{8.27}$$

At $O(\delta^4)$, equation (8.23) gives

$$0 = -2R_0^2R_4 - 3R_0R_2^2 - R_0^3\frac{\partial\phi_2}{\partial X} - R_0^2R_2\frac{\partial\phi_0}{\partial X} - \frac{R_0^3}{2}\left(\frac{\partial\phi_0}{\partial X}\right)^2 - 2d\frac{\partial\phi_0}{\partial X} + \frac{\partial^2 R_2}{\partial X^2}, \tag{8.28}$$

which can be reduced to

$$R_0^2\left(2R_4 + R_0\frac{\partial\phi_2}{\partial X}\right) = -\frac{3}{4}R_0^3\left(\frac{\partial\phi_0}{\partial X}\right)^2 - 2d\frac{\partial\phi_0}{\partial X} - \frac{R_0}{2}\frac{\partial^3\phi_0}{\partial X^3}, \tag{8.29}$$

using equation (8.27). In turn, equation (8.24) gives

$$R_0\frac{\partial\phi_0}{\partial T} = R_2\frac{\partial^2\phi_0}{\partial X^2} + R_0\frac{\partial^2\phi_2}{\partial X^2} + 2\frac{\partial R_4}{\partial X} + 2\frac{\partial R_2}{\partial X}\frac{\partial\phi_0}{\partial X}, \tag{8.30}$$

and combining this with equations (8.27) and (8.29) leads to the following nonlinear phase equation for ϕ_0

$$\frac{\partial\phi_0}{\partial T} = -\frac{d}{q_0^2}\frac{\partial^2\phi_0}{\partial X^2} - \frac{1}{4q_0^2}\frac{\partial^4\phi_0}{\partial X^4} - \frac{3}{2}\frac{\partial}{\partial X}\left(\frac{\partial\phi_0}{\partial X}\right)^2, \tag{8.31}$$

where $q_0^2 = R_0^2/2$. The evolution of the phase is therefore diffusive in character.

The nonlinear phase equation may also be determined solely from the linear growth rate of phase disturbances, using Kuramoto's method (Kuramoto, 1984a). We have already determined the appropriate growth rate, σ_2, which is given by equation (8.12) with $l = 0$. So to fourth order in k we have

$$\sigma = -k^2\left(1 - \frac{2q^2}{R_0^2}\right) - \frac{2q^4k^4}{R_0^6}, \tag{8.32}$$

which corresponds to the linear phase equation

$$\frac{\partial\phi}{\partial T} = \left(1 - \frac{2q^2}{R_0^2}\right)\frac{\partial^2\phi}{\partial X^2} - \frac{2q^4}{R_0^6}\frac{\partial^4\phi}{\partial X^4} \tag{8.33}$$

The phase equation must be invariant under the symmetries of the system, namely translation symmetry $\phi \to \phi + \widetilde{\phi}$, where $\widetilde{\phi}$ is a constant, and reflection in x ($X \to -X, \phi \to -\phi$). The lowest order nonlinear term permitted by these symmetries is

$$\frac{\partial}{\partial X}\left(\frac{\partial\phi}{\partial X}\right)^2, \tag{8.34}$$

so if the coefficient of this term is nonzero, the leading-order weakly nonlinear phase equation should take the form

$$\frac{\partial\phi}{\partial T} = D_\parallel(q)\frac{\partial^2\phi}{\partial X^2} + h(q)\frac{\partial^4\phi}{\partial X^4} + g(q)\frac{\partial}{\partial X}\left(\frac{\partial\phi}{\partial X}\right)^2, \tag{8.35}$$

where

$$D_\parallel(q) = 1 - \frac{2q^2}{R_0^2} = \frac{\mu - 3q^2}{\mu - q^2}, \tag{8.36}$$

$$h(q) = -\frac{2q^4}{R_0^6} = -\frac{2q^4}{(\mu - q^2)^3} \tag{8.37}$$

Now we can consider the change of variables $\phi = \widehat{\phi} + \Delta q X$, where $|\Delta q| \ll 1$. This leads to a new phase-diffusion equation

$$\frac{\partial \widehat{\phi}}{\partial T} = [D_{\parallel}(q) + 2g(q)\Delta q]\frac{\partial^2 \widehat{\phi}}{\partial X^2} + [h(q) + O(\Delta q)]\frac{\partial^4 \widehat{\phi}}{\partial X^4} + [g(q) + O(\Delta q)]\frac{\partial}{\partial X}\left(\frac{\partial \widehat{\phi}}{\partial X}\right)^2 \tag{8.38}$$

The $O(\Delta q)$ terms in the coefficients of $\partial^4\widehat{\phi}/\partial X^4$ and $(\partial/\partial X)(\partial\widehat{\phi}/\partial X)^2$ come from the transformation of higher-order nonlinear terms in the phase equation, just as the $O(\Delta q)$ term in the coefficient of $\partial^2\widehat{\phi}/\partial X^2$ comes from the lowest-order nonlinear term. The change of variables we have made is the same as changing q by a small amount $q\Delta q$. Performing this equivalent transformation $q \to q(1 + \Delta q)$ in equation (8.35), gives another phase equation for $\widehat{\phi}$

$$\frac{\partial \widehat{\phi}}{\partial T} = \left(D_{\parallel}(q) + \frac{\mathrm{d}D_{\parallel}}{\mathrm{d}q}q\Delta q\right)\frac{\partial^2 \widehat{\phi}}{\partial X^2} + \left(h(q) + \frac{\mathrm{d}h}{\mathrm{d}q}q\Delta q\right)\frac{\partial^4 \widehat{\phi}}{\partial X^4} + \left(g(q) + \frac{\mathrm{d}g}{\mathrm{d}q}q\Delta q\right)\frac{\partial}{\partial X}\left(\frac{\partial \widehat{\phi}}{\partial X}\right)^2 + O(\Delta q^2) \tag{8.39}$$

Equating the coefficients of $\partial^2\widehat{\phi}/\partial X^2$ in equations (8.38) and (8.39), we find that

$$2g(q)\Delta q = \frac{\mathrm{d}D_{\parallel}}{\mathrm{d}q}q\Delta q, \tag{8.40}$$

and so

$$g(q) = \frac{1}{2}q\frac{\mathrm{d}D_{\parallel}}{\mathrm{d}q} = -\frac{2q^2\mu}{(\mu - q^2)^2} \tag{8.41}$$

Near the bifurcation point, we have $\mu = 3q^2 - 2d\delta^2$, which to leading order gives $g = -3/2$, agreeing with the coefficient found by the previous method. The coefficients of the linear terms close to the bifurcation also agree with those in equation (8.31) if the rescaling used there is taken into account. The $O(\Delta q)$ contributions to the coefficients of $\partial^4\widehat{\phi}/\partial X^4$ and $\partial/\partial X(\partial\widehat{\phi}/\partial X)^2$, unspecified in equation (8.38), are equal to those in equation (8.39).

The calculation above shows that the nonlinear term introduces a space-varying part to the effective diffusion coefficient. To make this more explicit, we can rewrite the phase-diffusion equation (8.31) in the form

$$\frac{\partial \phi_0}{\partial T} = -\left(\frac{d}{q_0^2} + 3\frac{\partial \phi_0}{\partial X}\right)\frac{\partial^2 \phi_0}{\partial X^2} - \frac{1}{4q_0^2}\frac{\partial^4 \phi_0}{\partial X^4} \tag{8.42}$$

The Eckhaus instability is subcritical for systems governed by the Newell–Whitehead–Segel equation (8.2); this may be deduced by considering a solution of equation (8.42) in the form

$$\phi_0 = A(T)\sin kX + B(T)\sin 2kX + \cdots \tag{8.43}$$

Substituting this into equation (8.42) and equating coefficients of $\sin kX$ and $\sin 2kX$ on either side gives

$$\frac{\mathrm{d}A}{\mathrm{d}T} = \frac{k^2}{q_0^2}(d - d_c)A + 3ABk^3, \tag{8.44}$$

$$\frac{\mathrm{d}B}{\mathrm{d}T} = \frac{4k^2}{q_0^2}(d - 4d_c)B + \frac{3}{2}A^2k^3, \tag{8.45}$$

where $d_c = k^2/4$. The phase mode is linearly unstable for $d > d_c$, since the contribution $A(T)\sin kX$ starts to grow. Looking for stationary solutions to equations (8.44) and (8.45) gives

$$A^2 = \frac{2(d - d_c)(d - 4d_c)}{9q_0^4 d_c}, \tag{8.46}$$

if $Ak^2 \neq 0$. Then A^2 is positive if $d < d_c$ or $d > 4d_c$. If we are close to the bifurcation, we must have $d \simeq d_c$, and so steady bifurcated solutions exist close to the bifurcation point only for $d < d_c$. In the range $d < d_c$, the system is linearly stable, so this is a subcritical pitchfork bifurcation for each k. It is important to bear in mind that the sub- or supercriticality of the bifurcation depends on the relationship between the linear and nonlinear terms in the phase equation, which are derived from the governing equation, in this case the Newell–Whitehead–Segel equation (8.2). It is possible to have a supercritical Eckhaus instability in a system governed by different equations, for example in certain parameter régimes when quintic and space-derivative cubic terms are included in the envelope equation (see Hoyle, 1998a, for further details).

In the current situation the bifurcation is subcritical, so a roll pattern undergoing the Eckhaus instability does not saturate in the bifurcated state. Instead, the pattern breaks down creating defects, eliminating or adding more rolls to alter the wavelength until it lies in the stable band. In one-dimensional systems where there is no dependence on Y and the Newell–Whitehead–Segel equation (8.2) is replaced by the Ginzburg–Landau equation (7.48), the defects exist only for an instant while a pair of rolls is created or eliminated (Figure 8.2). On the other hand, in the simulation of the Newell–Whitehead–Segel equation shown in Figure 8.3 there is some variation in the y direction, owing to a small random perturbation of the initial

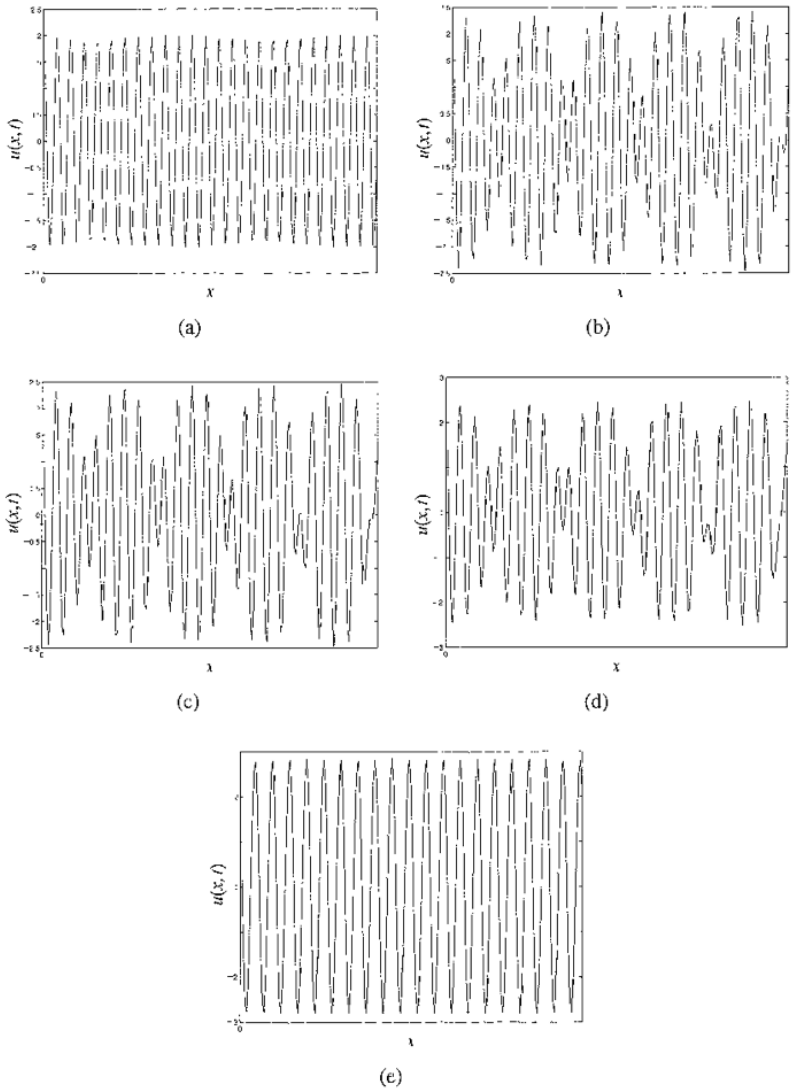
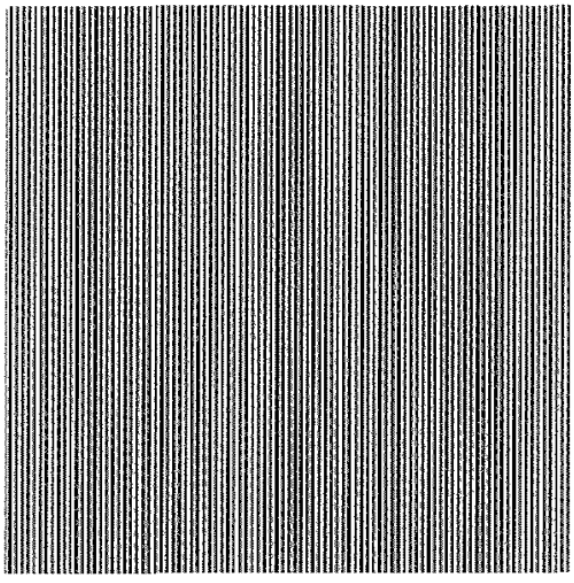
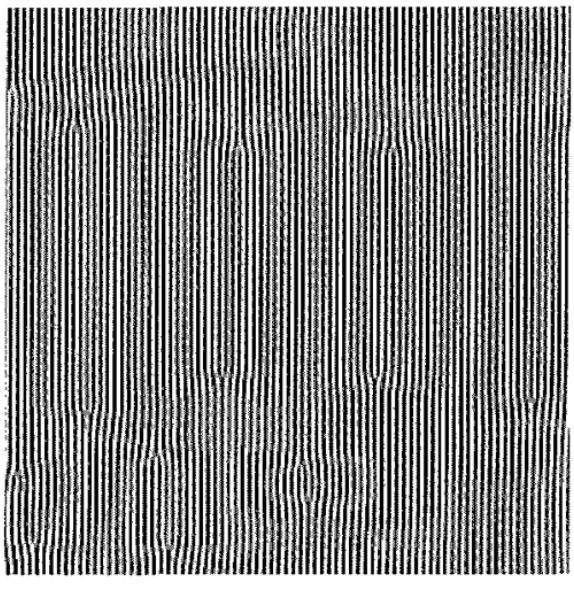


Fig. 8.2 The Eckhaus instability in the Ginzburg–Landau equation (7.48) with periodic boundary conditions. The pictures show plots of $u(x,t) = A(X,T)e^{ix} + c.c.$, where $\epsilon \equiv X/x = 0.19$. Only a quarter of the full simulation domain is shown. (a) The initial state at time $T = 0$ is a uniform roll solution with $q = 1.0$ and $\mu = 2.0$ in the Eckhaus-unstable region where the pattern wavelength is too short. A small random perturbation has been added so that the instability will proceed. The sequence of plots at times (b) $T = 16.7$, (c) $T = 16.9$ and (d) $T = 17.1$ shows the instantaneous formation of a defect ($|A| = 0$) that removes a roll pair at the righthand edge of the plot domain. By time $T = 26.7$ the pattern, shown in (e), has settled down to a new uniform roll solution at slightly longer wavelength.

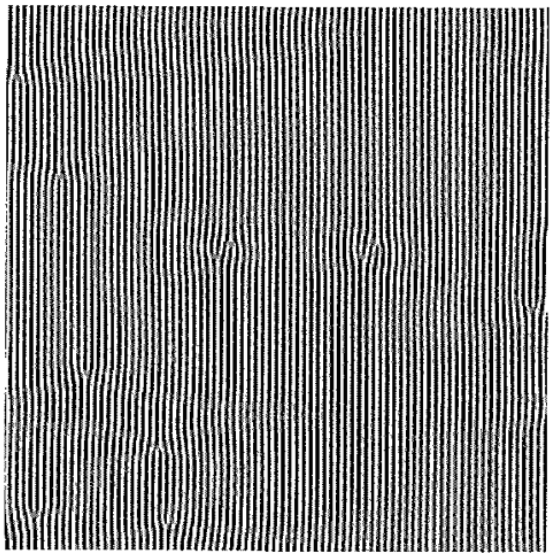


(a)

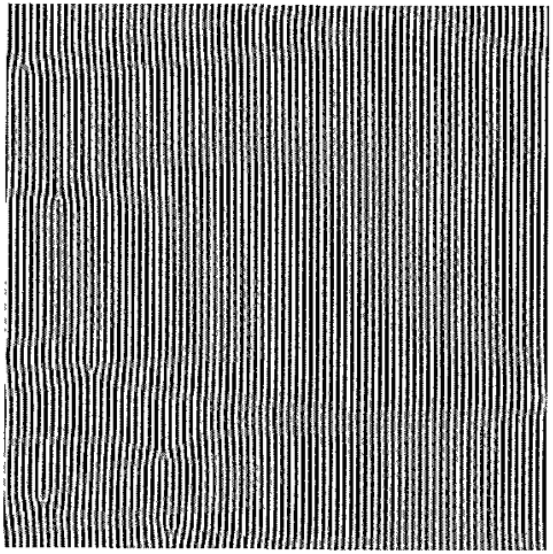


(b)

Fig. 8 3 The Eckhaus instability in the Newell–Whitehead–Segel equation (8 2) with periodic boundary conditions The pictures show greyscale plots of $u(x, y, t) = A(X, Y, T)e^{ix} + c.c$, where $\epsilon \equiv X/x = 0\ 19$ (a) The initial state at time $T = 0$ is once again a uniform roll solution with $q = 1.0$ and $\mu = 2\ 0$ in the Eckhaus-unstable region where the pattern wavelength is too short, with an additional small random perturbation Defects form and move through the pattern removing rolls and thus leading to an increase in the wavelength The progress of the instability is shown at times (b) $T = 428$, (c) $T = 855$ and (d) $T = 1425$



(c)



(d)

Fig. 8.3. (*cont.*)

state, and so the simple one-dimensional character of the Eckhaus instability is lost. The defects in this case persist for some time, moving through the pattern until they meet in pairs and annihilate each other (Figure 8.3). This is typical of real two-dimensional situations.

Another perspective on the creation of defects comes from looking at the evolution of the perturbation wavenumber. From the phase equation (8.42) we can

deduce the equation

$$\frac{\partial k}{\partial T} = \frac{\partial}{\partial X}\left\{\left(-\frac{d}{q_0^2}\frac{\partial k}{\partial X} + \frac{3}{2}k^2\right) - \frac{1}{4q_0^2}\frac{\partial^3 k}{\partial X^3}\right\} \tag{8.47}$$

for the evolution of the perturbation wavenumber $k = \partial\phi_0/\partial X$. Integrating this over the whole domain gives

$$\frac{\mathrm{d}}{\mathrm{d}T}\int_{-\infty}^{+\infty} k\,\mathrm{d}x = 0, \tag{8.48}$$

with appropriate boundary conditions at $\pm\infty$, such as the perturbation being zero there. This says that the mean wavenumber is conserved in time, and so the pattern cannot change its wavelength through a continuous, smooth deformation. We know that wavenumber changes do come about during the Eckhaus instability, and so we expect defects to arise in the pattern, where pairs of rolls are created or annihilated discontinuously, and where rapid spatial variations of the amplitude lead to the breakdown of the envelope formalism, so that the phase equation (8.42) is no longer valid.

Further information on the properties of the Eckhaus instability can be found in the article by Kramer and Zimmerman (1985).

8.2.1 Compression-dilatation waves

The Eckhaus instability is a compression and dilatation instability. We can see what happens when a stationary stripe pattern is compressed or dilated by writing $A = Re^{i\theta}$ and looking at the angular momentum equation

$$h = R^2\frac{\mathrm{d}\theta}{\mathrm{d}X}, \tag{8.49}$$

where h is constant in space, which is equation (7.74) from the previous chapter. If the pattern is compressed or dilated, there are changes in the wavenumber, so we expect $\mathrm{d}\theta/\mathrm{d}X \neq 0$ and hence $h \neq 0$. The phase portrait in Figure 7.4a ($\mu > 0$, $h \neq 0$) is relevant here. The homoclinic orbits marked in bold correspond to a localised compression ($h > 0$) or dilatation ($h < 0$) of the roll pattern (Figure 8.4). The system is symmetric in R, so we need only consider $R > 0$. In the compressed or dilated region the amplitude decreases as we move along the homoclinic orbit to the left of the fixed point in $R > 0$, and so according to conservation of angular momentum, h, the modulus of the perturbation wavenumber $|\mathrm{d}\theta/\mathrm{d}X|$ must increase. If h is negative then $\mathrm{d}\theta/\mathrm{d}X$ becomes more negative, the overall wavenumber decreases, and so there is dilatation, whereas if h is positive, there is compression because $\mathrm{d}\theta/\mathrm{d}X$ increases. Fauve (1991, 1998)

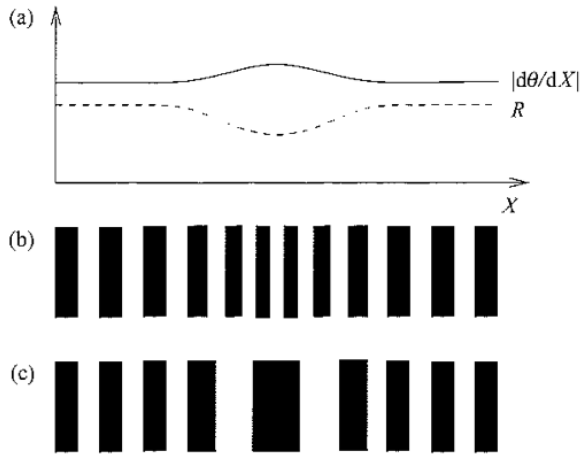


Fig 8.4 (a) Sketches of the profiles of R and $|\mathrm{d}\theta/\mathrm{d}X|$ for stationary solutions describing a localised (b) compression ($h > 0$, $\mathrm{d}\theta/\mathrm{d}X > 0$) or (c) dilatation ($h < 0$, $\mathrm{d}\theta/\mathrm{d}X < 0$) of the roll pattern. These compression/dilatation solitary waves correspond to the $R > 0$ homoclinic orbit marked in bold in Figure 7 4a

gives further information on this solitary compression/dilatation wave solution and its stability. In particular he comments that this wave represents the critical localised perturbation that gives rise to the finite amplitude Eckhaus instability in the subcritical region where an off-critical uniform roll pattern is linearly stable

More generally, the conservation of h shows that if there is a high degree of compression or dilatation in a region, then the amplitude R must be very small there. If the amplitude vanishes then the phase, θ, is undetermined, and so a pair of rolls can be created or annihilated This defect-forming process cannot accurately be described by the envelope formalism, however, as it requires the amplitude to vary over lengthscales comparable with that of the carrier wave

8.2.2 Side bands

The Eckhaus instability is sometimes called a **side-band instability**, because in a band of unstable modes centred on the wavenumber q, a side mode $q + k$ interacts with the second harmonic $2q$ to give the mode $q - k$ on the other side of the band, since $2q - (q + k) = q - k$ This resonance causes the side modes to be amplified, leading to the instability. To see this we can analyse the instability in a slightly different way. We write

$$A = (\mu - q^2)^{1/2} e^{iqX}(1 + a(I)e^{ikX} + \bar{b}(I)e^{-ikX}), \tag{8.50}$$

where $|a|, |b| \ll 1$. This a roll/stripe pattern perturbed by side modes $q + k$ and $q - k$. Substituting into the Ginzburg–Landau equation,

$$\frac{\partial A}{\partial T} = \mu A - |A|^2 A + \frac{\partial^2 A}{\partial X^2}, \tag{8.51}$$

linearising in a and b, and equating the coefficients of $e^{i(q+k)X}$ and $e^{i(q-k)X}$ on each side of the equation gives

$$\frac{\mathrm{d}a}{\mathrm{d}T} = -2qka - k^2 a - (\mu - q^2)(a + b), \tag{8.52}$$

$$\frac{\mathrm{d}b}{\mathrm{d}T} = 2qkb - k^2 b - (\mu - q^2)(a + b). \tag{8.53}$$

The presence of b in the equation for $\mathrm{d}a/\mathrm{d}T$ comes from the interaction in the term $|A|^2 A$ of the second harmonic, e^{2iqX}, and the complex conjugate, $e^{-i(q+k)X}$, of the side mode corresponding to a. The presence of a in the $\mathrm{d}b/\mathrm{d}T$ equation can be explained analogously. These are the resonant terms leading to the Eckhaus instability.

If a and b are proportional to $e^{\sigma T}$, then by using equations (8.52) and (8.53) it is straightforward to show that the growth rate eigenvalues, σ, satisfy

$$\sigma^2 + 2\sigma(k^2 + \mu - q^2) + k^4 + 2\mu k^2 - 6q^2 k^2 = 0. \tag{8.54}$$

There are two solutions for σ:

$$\sigma_1 = -2(\mu - q^2) + O(k^2), \tag{8.55}$$

$$\sigma_2 = -\frac{(\mu - 3q^2)}{(\mu - q^2)} k^2 + O(k^4). \tag{8.56}$$

The first we recognise from equation (8.11) as the rate of decay of amplitude perturbations, and the second from equation (8.21) as the growth rate governing phase perturbations – in both cases substituting $R_0^2 = \mu - q^2$. In the first case we need only consider the $O(1)$ terms in the evolution equations for a and b, which can then be rearranged to give

$$\frac{\mathrm{d}}{\mathrm{d}T}(a + b) = -2(\mu - q^2)(a + b), \tag{8.57}$$

$$\frac{\mathrm{d}}{\mathrm{d}T}(a - b) = 0. \tag{8.58}$$

We can now identify $(a + b)$ as representing the amplitude perturbation, while $(a - b)$ will represent the phase perturbation.

Substituting σ_2 into the equations for a and b shows that the phase perturbation eigenvector satisfies

$$(a+b) = -\frac{qk}{\mu - q^2}(a-b) + O(k^2). \tag{8.59}$$

At leading order in k then we have $(a+b) = 0$, and so phase perturbations are indeed represented by $(a-b)$

8.2.3 The effect of finite domain size

It turns out that in a finite domain the Eckhaus instability does not work quite as expected This was first pointed out by Tuckerman and Barkley (1990). The growth rate eigenvalue solutions to equation (8.54), σ_1 and σ_2, can be written in the form

$$\sigma_{1,2} = -(\mu - q^2 + k^2) \mp \sqrt{(\mu - q^2)^2 + 4q^2k^2} \tag{8.60}$$

As we have seen, we must have $\sigma_1 < 0$, since $\mu > q^2$, so the original roll solution can never be unstable to the corresponding amplitude eigenmode On the other hand the phase eigenmode can grow if σ_2 is positive, which happens when

$$(\mu - q^2)^2 + 4q^2k^2 > (\mu - q^2 + k^2)^2, \tag{8.61}$$

or equivalently

$$\mu < 3q^2 - \frac{k^2}{2} \tag{8.62}$$

Now if the roll solution is to be stable to perturbations at all wavenumbers k, then we must have

$$\mu > 3q^2 - \frac{k^2}{2} \tag{8.63}$$

for all admissible k If the domain is infinite, we can have perturbations of any wavenumber, k, so the stability boundary is given by $\mu = 3q^2$, the usual Eckhaus boundary If, however, the domain is finite then both q and k are quantised so that the solution $u(x,t)$ fits into the box, so the Eckhaus boundary will be shifted downwards by an amount $k_b^2/2$, where k_b^2 is the smallest permitted k^2.

To see how this works in detail, assume that the solution is periodic on a domain of length L. Then in order for the roll solution to fit into the domain the wavenumber, $1 + \epsilon q$, must satisfy

$$1 + \epsilon q = \frac{2\pi n}{L}, \tag{8.64}$$

where n is an integer, and $X = \epsilon x$ with $|\epsilon| \ll 1$ as usual. Similarly, if the perturbation is to fit then the perturbation wavenumber, k, must satisfy

$$1 + \epsilon(q + k) = \frac{2\pi m}{L}, \tag{8.65}$$

where $m \neq n$ is also an integer. We can also express the domain length in the form

$$L = 2\pi(N + l), \tag{8.66}$$

where N is an integer, and $|l| \leq 1/2$. Now the neutral stability curve and Eckhaus stability boundaries can be rewritten

$$\mu_N = \left(\frac{2\pi}{\epsilon L}\right)^2 (n - N - l)^2, \tag{8.67}$$

$$\mu_E = \left(\frac{2\pi}{\epsilon L}\right)^2 \left(3(n - N - l)^2 - \frac{(m - n)^2}{2}\right) \tag{8.68}$$

respectively. The roll solution corresponding to a particular choice of n will be stable to all Eckhaus perturbations if $\mu_E < \mu_N$ for all integers m, in other words if

$$(n - N - l)^2 < \frac{(m - n)^2}{4} \tag{8.69}$$

holds for all m. Recall that N is set by the length, L, of the domain. The lefthand side will be smallest for $n = N$, and since $m \neq n$, the righthand side is smallest for $|m - n| = 1$, so the mode $n = N$ will be stable to the Eckhaus instability if

$$l^2 < \frac{1}{4} \tag{8.70}$$

holds. Now l is defined such that $|l| \leq 1/2$. The special cases $|l| = 0$ and $|l| = 1/2$, when the domain length is equal to a whole number of half wavelengths, lead to the simultaneous bifurcation of the modes $\pm|q|$. We shall avoid such complications, and consider only the case $0 < |l| < 1/2$, for which inequality (8.70) holds and the roll solution is therefore stable to the Eckhaus instability. This is a somewhat surprising result! The roll wavenumber is not critical, since $q \neq 0$, so in an infinite domain, we would expect there to be a range of μ, given by $q^2 < \mu < 3q^2$, for which the pattern should be Eckhaus-unstable. However the quantisation of wavenumbers implied by the finite domain has shifted the Eckhaus stability curve below the neutral stability, or existence, curve in the neighbourhood of $q = 2\pi l/\epsilon L$, and the pattern is unexpectedly stable. For any other permitted value of q, corresponding to a value of n other than N, the lefthand side of inequality (8.69) is greater than $1/4$ and so the roll solution is always unstable for some value of μ. The arrangement of the neutral and Eckhaus stability curves is shown in Figure 8.5

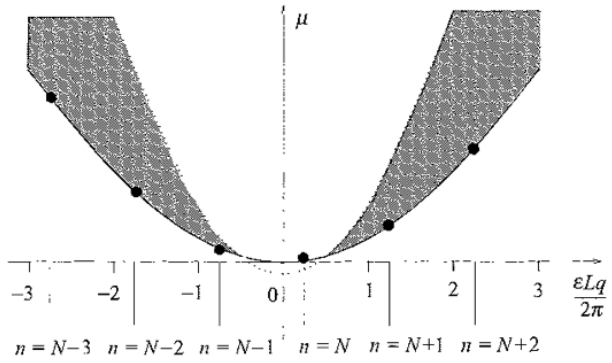


Fig 8.5 The neutral stability curve, $\mu = q^2$, (solid line) and Eckhaus instability boundary, $\mu = 3q^2 - 2\pi^2/\epsilon^2 L^2$, (dotted line) for a finite domain of length L and periodic boundary conditions. The Eckhaus-unstable region is shaded grey The large black dots represent the points at which the various permitted roll modes $q_n = 2\pi(n - N - l)/\epsilon L$ bifurcate from the zero solution. Rolls at wavenumber q_N are Eckhaus-stable, since the Eckhaus-stability boundary passes below the neutral stability curve at that point

The full bifurcation analysis of the Eckhaus instability in a finite domain can be found in Tuckerman and Barkley (1990) Free-slip boundary conditions $\partial u/\partial x = 0$ at $x = 0, L$ are treated along with the periodic case.

8.3 The zigzag instability

The zigzag instability arises from perturbations that vary only along the Y direction, parallel to the roll axes. Setting $k = 0$ in equation (8.12) gives the growth rate

$$\sigma_2 = -ql^2 - \frac{l^4}{4}, \tag{8.71}$$

for phase perturbations Rolls are therefore unstable to phase modes that vary along the roll axes when $q < 0$, for small enough $|l|$ ($l^2 < -4q$). This is the zigzag instability (Busse, 1962; Schlüter, Lortz & Busse, 1965; Newell & Whitehead, 1969; Busse & Whitehead, 1971), which creates undulations along the roll axes when the wavelength is too large, thereby reducing it. The unstable region is shown in Figure 8.1b.

The nonlinear phase-diffusion equation relevant to the zigzag instability is found using a weakly nonlinear analysis close to onset. The linear part of the equation is given in equation (8 18) with $\partial/\partial X \equiv 0$, and the leading order nonlinear term can be discovered using symmetry constraints. The phase equation must

obey equivariance under translation in x

$$\phi \to \phi + \widetilde{\phi}, \tag{8.72}$$

where $\widetilde{\phi}$ is a constant, x-reflection

$$X \to -X, \ \phi \to -\phi, \tag{8.73}$$

and y-reflection

$$Y \to -Y \tag{8.74}$$

The leading order nonlinear term consistent with these symmetries is

$$\frac{\partial}{\partial Y}\left(\frac{\partial \phi}{\partial Y}\right)^3, \tag{8.75}$$

so to fourth order in the Y-derivative, the phase equation can be written

$$\frac{\partial \phi}{\partial T} = q\frac{\partial^2 \phi}{\partial Y^2} - \frac{1}{4}\frac{\partial^4 \phi}{\partial Y^4} + g\frac{\partial}{\partial Y}\left(\frac{\partial \phi}{\partial Y}\right)^3 \tag{8.76}$$

Close to onset, we can rescale $q \to \delta^2 q$. In order that all terms in equation (8.76) contribute at the same order, we then scale

$$\frac{\partial}{\partial Y} \to \delta\frac{\partial}{\partial Y}, \quad \frac{\partial}{\partial T} \to \delta^4\frac{\partial}{\partial T} \tag{8.77}$$

We do not rescale ϕ since it must be $O(1)$ if all terms in equation (8.76) are to be of the same order. In contrast to the case of the Eckhaus instability, where $\phi \sim \delta$, the phase here is not small. For this analysis we define ϕ such that $A = Re^{i(qX+\phi)}$, so the phase differs from that used in the previous section by a factor of q; this is because we are interested in the behaviour in the vicinity of the zigzag instability boundary $q = 0$ and we do not want to restrict attention to small ϕ.

If we expand the amplitude and phase variables so that

$$R = R_0 + \delta^2 R_2 + \delta^4 R_4 + \cdots, \tag{8.78}$$

$$\phi = \phi_0 + \delta^2\phi_2 + \delta^4\phi_4 + \cdots, \tag{8.79}$$

and perform a weakly nonlinear analysis in the same manner as we did for the Eckhaus instability, we arrive at the nonlinear phase-diffusion equation

$$\frac{\partial \phi_0}{\partial T} = q\frac{\partial^2 \phi_0}{\partial Y^2} - \frac{1}{4}\frac{\partial^4 \phi_0}{\partial Y^4} + \frac{1}{2}\frac{\partial}{\partial Y}\left(\frac{\partial \phi_0}{\partial Y}\right)^3. \tag{8.80}$$

The zigzag instability differs from the Eckhaus in that it is supercritical rather than subcritical for systems governed by the Newell–Whitehead–Segel equation

We can see this by rewriting equation (8.80) in the form

$$\frac{\partial \phi}{\partial T} = \left(q + \frac{3}{2} \left(\frac{\partial \phi}{\partial Y} \right)^2 \right) \frac{\partial^2 \phi}{\partial Y^2} - \frac{1}{4} \frac{\partial^4 \phi}{\partial Y^4} \tag{8.81}$$

The nonlinear term can be thought of as contributing a positive space-varying component to the diffusion coefficient. In the region of linear instability, we have $q < 0$, while the nonlinear term in the diffusion coefficient is positive, and therefore stabilising, and so quenches the instability. The bifurcation is supercritical, and the roll pattern saturates in the bifurcated state. (Again, the relationship between the linear and nonlinear terms in the phase equation determines whether or not the instability is supercritical, and for different choices of governing equations it can be subcritical – see Hoyle, 1998a). The rolls become wavy owing to nonlinear effects, as shown in Figure 8.6. The zigzag instability occurs when rolls have too long a wavelength; bending them produces a shorter wavelength, as illustrated in Figure 8.7, and therefore stabilises the pattern. The new pattern grows to fill the available space, and eventually the rolls reconnect at the bends to form patches of oblique rolls that meet at **grain boundaries**. Point defects are not, however, created during the zigzag instability, and grain boundaries form only at a late stage in the the nonlinear evolution.

The coefficient of the nonlinear term in the zigzag phase equation can also be determined from the linear diffusion coefficients using a calculation similar to Kuramoto's for the Eckhaus instability. Starting from equation (8.76) we write

$$\frac{\partial \phi}{\partial T} = D_\perp(q) \frac{\partial^2 \phi}{\partial Y^2} - \frac{1}{4} \frac{\partial^4 \phi}{\partial Y^4} + g(q) \frac{\partial}{\partial Y} \left(\frac{\partial \phi}{\partial Y} \right)^3, \tag{8.82}$$

where $D_\perp(q) = q$. The solution $\phi = pY$, corresponds to tilted rolls given by

$$u(x, y, t) = A e^{i(\lambda + qX + pY)} + c.c. + h.o.t., \tag{8.83}$$

$$= A e^{i(x + \epsilon q x + \epsilon^{\frac{1}{2}} py)} + c.c. + h.o.t., \tag{8.84}$$

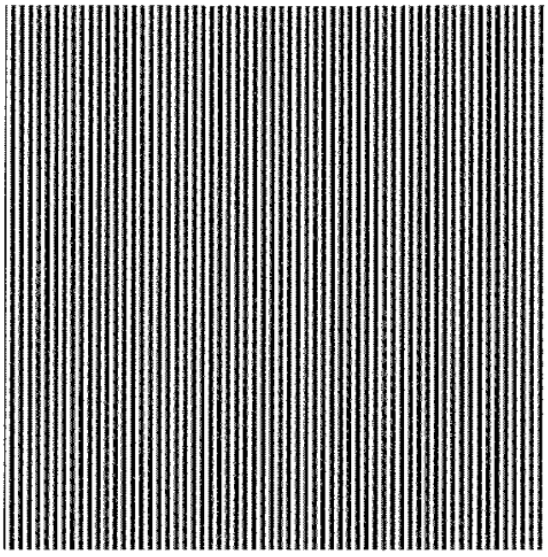
$$= A e^{i\sqrt{(1+\epsilon q)^2 + \epsilon p^2}\tilde{x}} + c.c. + h.o.t., \tag{8.85}$$

$$= A e^{i(1 + \epsilon q + \epsilon p^2/2 + O(\epsilon^2))\tilde{x}} + c.c. + h.o.t., \tag{8.86}$$

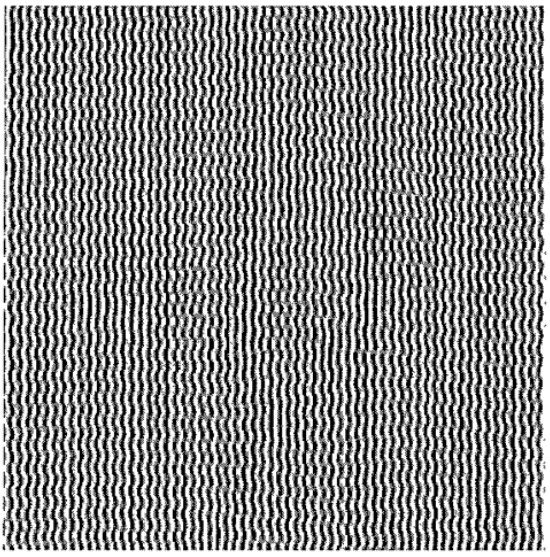
$$\approx A e^{i(\tilde{x} + (q + p^2/2)\tilde{X})} + c.c. + h.o.t., \tag{8.87}$$

where the $\tilde{x}$ direction is perpendicular to the axes of the tilted rolls (see Figure 8.8), and $\tilde{X} = \epsilon \tilde{x}$. So these tilted rolls have perturbation wavenumber q', given by

$$q' = q + \frac{p^2}{2} \tag{8.88}$$

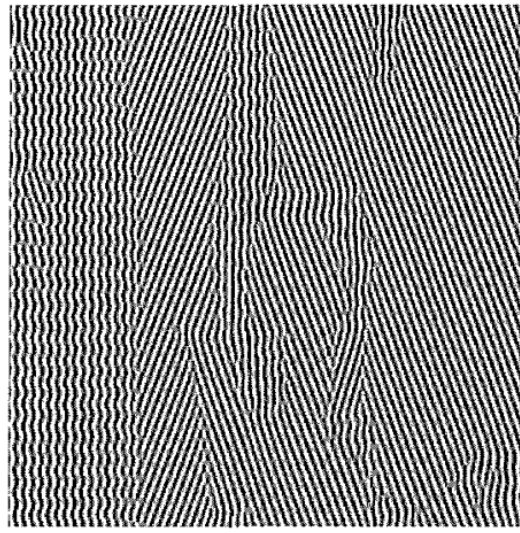


(a)

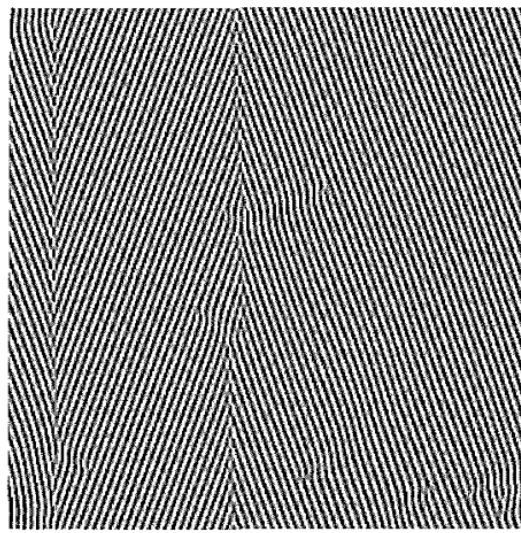


(b)

Fig. 8 6 The zigzag instability in the Newell–Whitehead–Segel equation (8 2) with periodic boundary conditions. The pictures show greyscale plots of $u(x, y, t) = A(X, Y, T)e^{ix} + c.c.$, where $\epsilon \equiv X/x = 0\ 20$ (a) The initial state at time $T = 0$ is a uniform roll solution with $q = -1\ 0$ and $\mu = 4.0$ in the zigzag unstable region where the pattern wavelength is too long A small random perturbation has been added to seed the instability. (b) At time $T = 16$, the rolls have become wavy (c) At time $T = 32$ the rolls have reconnected at the bends to form rolls oblique to the original pattern and at shorter wavelength (d) At time $T = 158$ the pattern consists of patches of oblique rolls that meet at lines called grain boundaries



(c)



(d)

Fig 8 6. (*cont*)

Let us assume that $p^2 \ll |q|$, and substitute $\phi = pY + \widehat{\phi}$ into equation (8 82) to give

$$\frac{\partial \widehat{\phi}}{\partial T} = (D_\perp(q) + 3g(q)p^2)\frac{\partial^2 \widehat{\phi}}{\partial Y^2} - \left(\frac{1}{4} + O(p^2)\right)\frac{\partial^4 \widehat{\phi}}{\partial Y^4}$$

$$+ (g(q) + O(p^2))\frac{\partial}{\partial Y}\left(\frac{\partial \widehat{\phi}}{\partial Y}\right)^3, \tag{8.89}$$

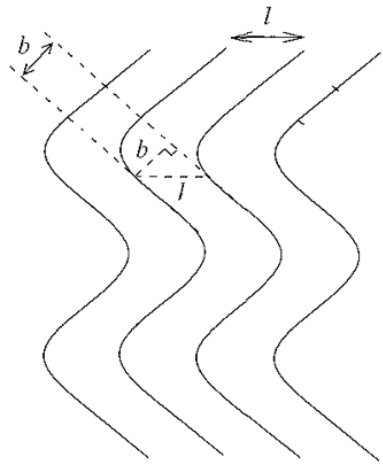


Fig 8 7. Bending a roll pattern produces a shorter wavelength The new wavelength, b, must be less than the original wavelength, l, by Pythagoras' theorem. The zigzag instability uses this mechanism to evolve from an unstable pattern at too long a wavelength into a pattern with a shorter wavelength in the stable band

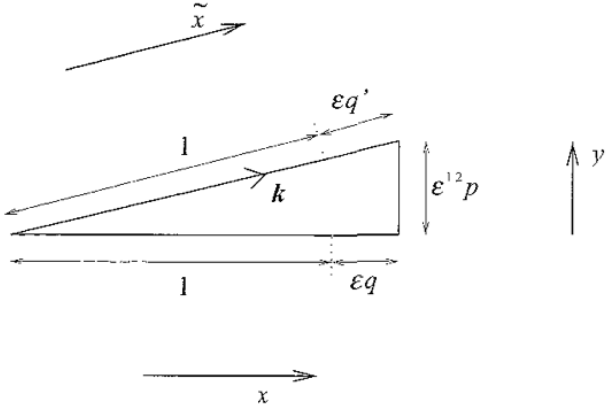


Fig. 8 8. Variables associated with the tilted roll solution $\phi = pY$ to equation (8 82). The wavevector of the tilted rolls is $\boldsymbol{k}$.

where the $O(p^2)$ terms in the coefficients of $\partial^4\widehat{\phi}/\partial Y^4$ and $(\partial/\partial Y)(\partial\widehat{\phi}/\partial Y)^3$ come from higher-order nonlinear terms in the phase equation.

The new phase, $\widehat{\phi}$ must also satisfy the phase-diffusion equation for tilted rolls at wavenumber $q' = q + p^2/2$. We want to compare the coefficient of the diffusion term $\partial^2\widehat{\phi}/\partial Y^2$ with that in equation (8.89), but since the rolls are tilted we will have to retain both the terms in $\partial^2\widehat{\phi}/\partial\widetilde{X}^2$ and $\partial^2\widehat{\phi}/\partial\widetilde{Y}^2$ in order to compute the coefficient Thus we have

$$\frac{\partial\widehat{\phi}}{\partial T} = D_{\parallel}\left(q + \frac{p^2}{2}\right)\frac{\partial^2\widehat{\phi}}{\partial\widetilde{X}^2} + D_{\perp}\left(q + \frac{p^2}{2}\right)\frac{\partial^2\widehat{\phi}}{\partial\widetilde{Y}^2} + h.o.t\,, \qquad (8\,90)$$

where $D_\parallel(q) = 1 - 2q^2/(\mu - q^2)$ as before This is the truncation of equation (8 18) at leading order

The tilde variables are related to the plain variables according to

$$\widetilde{x} = \{x(1+\epsilon q) + \epsilon^{\frac{1}{2}} py\}\{(1+\epsilon q)^2 + \epsilon p^2\}^{-\frac{1}{2}}, \tag{8 91}$$

$$\widetilde{y} = \{y(1+\epsilon q) - \epsilon^{\frac{1}{2}} px\}\{(1+\epsilon q)^2 + \epsilon p^2\}^{-\frac{1}{2}} \tag{8 92}$$

Since $\widetilde{X} = \epsilon\widetilde{x}$ and $\widetilde{Y} = \epsilon^{1/2}\widetilde{y}$, we have

$$\widetilde{X} = X + O(\epsilon^{3/2}), \tag{8 93}$$

$$\widetilde{Y} = Y - pX + O(\epsilon^{3/2}), \tag{8 94}$$

and

$$\frac{\partial}{\partial \widetilde{X}} = \frac{\partial}{\partial X} + p\frac{\partial}{\partial Y} + O(\epsilon^{3/2}), \tag{8 95}$$

$$\frac{\partial}{\partial \widetilde{Y}} = \frac{\partial}{\partial Y} + O(\epsilon^{3/2}). \tag{8 96}$$

Now setting $\partial/\partial X = 0$ to match the conditions for the derivation of equation (8 89) and substituting into equation (8.90) gives

$$\frac{\partial \widehat{\phi}}{\partial T} = \left(D_\parallel(q)p^2 + D_\perp(q) + \frac{p^2}{2}\frac{\mathrm{d}D_\perp}{\mathrm{d}q}\right)\frac{\partial^2 \widehat{\phi}}{\partial Y^2} + h.o.t. \tag{8.97}$$

At the zigzag instability boundary, $q = 0$, the diffusion coefficient in equation (8 97) must be the same as that in equation (8 89), so the coefficient $g(0)$ of the nonlinear term is given by

$$g(0) = \frac{1}{3}D_\parallel(0) + \frac{1}{6}\frac{\mathrm{d}D_\perp}{\mathrm{d}q} = \frac{1}{3} + \frac{1}{6} = \frac{1}{2} \tag{8.98}$$

This agrees with the value given in equation (8 80), derived using a formal asymptotic expansion in powers of δ.

Manneville (1990) and Fauve (1991) both go through a calculation similar to the one just presented, but they neglect the $p\partial/\partial Y$ component of $\partial/\partial\widetilde{X}$, and so they come up with a different answer for g that does not agree with equation (8.80)

8.4 A general theory of phase dynamics

Phase dynamics can be studied directly using an appropriate formalism. This elegant approach was introduced by Pomeau and Manneville (1979), and has been widely used since (see, for example, Manneville 1990, Fauve 1991 and 1998, Chossat and Iooss, 1994) Unlike the methods that we have used so far it can be

applied to finite amplitude patterns, far from onset of the primary, pattern-forming instability.

Imagine that we have a stationary space-periodic solution $u_0(x)$ of the partial differential equation

$$\frac{\partial u(x, y, t)}{\partial t} = f(u(x, y, t), \mu), \tag{8.99}$$

where $x, y \in \mathbb{R}$, and f is a smooth nonlinear operator that includes spatial derivatives, but no time-dependence or explicit dependence on x or y. We shall assume that equation (8.99) holds on the infinite plane, and has Euclidean symmetry. Since there is symmetry under translation in x, $u_0(x + \phi)$ is also a stationary solution of the equation for all constant ϕ, so we have

$$f(u_0(x + \phi), \mu) = 0. \tag{8.100}$$

If now ϕ is time-dependent, so that

$$u(x, y, t) = u_0(x + \phi(t)), \tag{8.101}$$

then since f is independent of time and does not depend explicitly on x or y, we also have

$$f(u_0(x + \phi(t)), \mu) = 0 \tag{8.102}$$

Substituting the solution (8.101) into equation (8.99) gives

$$\frac{\mathrm{d}u_0}{\mathrm{d}x}\frac{\mathrm{d}\phi}{\mathrm{d}t} = f(u_0(x + \phi(t)), \mu) = 0, \tag{8.103}$$

so the translation, or phase, mode is neutral, and perturbations along the group orbit of translations, namely phase perturbations, take the form

$$\frac{\mathrm{d}u_0}{\mathrm{d}x}\phi \tag{8.104}$$

This is as expected from the general result at the end of Section 4.1.1 relating continuous symmetries, such as translations, to zero eigenvalues.

Differentiating equation (8.100) with respect to ϕ gives the useful result

$$\left.\frac{\delta f}{\delta u}\right|_{u = u_0} \left(\frac{\mathrm{d}u_0(x)}{\mathrm{d}x}\right) = 0, \tag{8.105}$$

where $\delta f/\delta u$ is the Fréchet, or variational, derivative of f with respect to u, defined by

$$\lim_{\epsilon' \to 0} \frac{1}{\epsilon'} \left\{ \int_{-L/2}^{L/2} \left(f(u(x) + \epsilon' h(x), \mu) - f(u(x), \mu) - \epsilon' \frac{\delta f}{\delta u} h(x) \right) \mathrm{d}x \right\} = 0, \tag{8.106}$$

for arbitrary $h(x)$ where L is the spatial period of $u_0(x)$.

Now assume the phase takes the more general form $\phi(X, Y, T)$, where X, Y and T are long scales in space and time, given by $X = \epsilon x$, $Y = \epsilon^p y$ and $T = \epsilon^r t$, where $|\epsilon| \ll 1$. So far all the phase equations we have looked at are diffusive, so that $r = 2$, but since we could have different behaviour we allow the power r to remain unspecified here. The exponent p is also left unspecified at this point. The solution is now

$$u(x, y, t) = u_0(x + \phi(X, Y, T)) + \epsilon^s u_1(x, X, Y, T) + \epsilon^{2s} u_2(x, X, Y, T) + \cdots, \tag{8.107}$$

where the higher order corrections, u_1, u_2 and so on, are necessary because $u_0(x + \phi(X, Y, T))$ is not an exact solution of equation (8.99) when ϕ is not constant in space. Again we allow the power, s, of ϵ multiplying u_1 to remain unspecified at present. Substituting the solution (8.107) into equation (8.99) and solving at every order in ϵ eventually leads to an evolution equation for the phase $\phi(X, Y, T)$. In this approach, although the phase gradients are small, because X and Y are long scales, neither the phase itself nor the stationary solution $u_0(x)$ need be, so the phase equation is valid for a wider range of solutions and perturbations than those we have derived so far.

If we want to consider only long wave perturbations of the original stationary solution $u_0(x)$ then we can simplify the form of the solution to

$$u(x, y, t) = u_0(x + \phi(X, Y, T)) + u_\perp(X, Y, T), \tag{8.108}$$

where $u_\perp(X, Y, T)$ represents perturbations transverse to the group orbit of translations, namely amplitude perturbations. We have just seen that constant phase perturbations have zero eigenvalue, so we can expect the phase $\phi(X, Y, T)$ to evolve slowly. In contrast, amplitude perturbations decay quickly if $u_0(x)$ is stable, so we can eliminate $u_\perp(X, Y, T)$ adiabatically to produce the phase equation for $\phi(X, Y, T)$. Since the equation must be equivariant under translations $\phi \to \phi + \tilde{\phi}$ for constant $\tilde{\phi}$, it can only contain derivatives of ϕ and so takes the form

$$\frac{\partial \phi}{\partial T} = g\left(\frac{\partial \phi}{\partial X}, \frac{\partial \phi}{\partial Y}, \frac{\partial^2 \phi}{\partial X^2}, \frac{\partial^2 \phi}{\partial X \partial Y}, \frac{\partial^2 \phi}{\partial Y^2}, \cdots\right). \tag{8.109}$$

One approach is simply to write down the leading order phase equation using the equivariance conditions. If on the other hand we want to derive the phase equation directly from the governing equation, we will have to start from equation (8.107) and perform an asymptotic expansion as outlined above. To see how this works, it will be helpful to use an example, so we will re-examine the zigzag instability using this approach.

8.4.1 The zigzag instability for finite amplitude rolls in the Swift–Hohenberg equation

We need a governing partial differential equation to start from: as an example, we will use the Swift–Hohenberg equation, but the same methods can be applied to other governing equations. The Swift–Hohenberg equation (7.1) is

$$\frac{\partial u}{\partial t} = [\mu - (\nabla^2 + 1)^2]u - u^3, \tag{8.110}$$

where the critical wavenumber k_c has been set to 1. We now look for solutions of the form

$$u(x, y, t) = u_0(x + \phi(Y, T)) + \epsilon^2 u_1(x, Y, T) + \epsilon^4 u_2(x, Y, T) + \quad , \tag{8.111}$$

where $u_0(x)$ is a fully nonlinear stripe (roll) solution, in other words a spatially periodic stationary solution of equation (8.110), and where $T = \epsilon^2 t$ and $Y = \epsilon y$. The phase, $\phi(Y, T)$, depends only on Y and not on X because the zigzag instability is characterised by perturbations along the roll axes. The scalings in ϵ of u_1, u_2, T and Y are chosen to give a phase-diffusion equation at leading order. Note that the scaling for Y is different from that used close to onset ($Y = \epsilon^{1/2} y$) in the Newell–Whitehead–Segel equation. If we didn't know the appropriate scalings we could allow the powers of ϵ to remain unspecified initially, and we would fix them during the course of the calculation to bring in the temporal and spatial derivatives in the phase equation at the same order.

The spatial derivatives are given by

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \epsilon^2 \frac{\partial^2}{\partial Y^2}, \tag{8.112}$$

while the time derivative is given by $\partial/\partial t = \epsilon^2 \partial/\partial T$, and so the Swift–Hohenberg equation becomes

$$\epsilon^2 \frac{\partial u}{\partial T} = \mu u - \left(\frac{\partial^2}{\partial x^2} + 1\right)^2 u - 2\epsilon^2 \frac{\partial^2}{\partial Y^2}\left(\frac{\partial^2}{\partial x^2} + 1\right) u - \epsilon^4 \frac{\partial^4 u}{\partial Y^4} - u^3 \tag{8.113}$$

Note that μ is not small here, in contrast to the close to onset case. Now substituting for u from equation (8.111) and equating powers of ϵ on both sides we find at $O(1)$

$$\mu u_0(x + \phi) - \left(\frac{\partial^2}{\partial x^2} + 1\right)^2 u_0(x + \phi) - u_0(x + \phi)^3 = 0. \tag{8.114}$$

This equation does indeed hold, since $u_0(x)$ is a stationary solution of equation (8.110) and $\phi(Y, T)$ has no x-dependence.

At $O(\epsilon^2)$ we have

$$\frac{du_0}{dx}\frac{\partial\phi}{\partial T} = \mathcal{L}_0 u_1 - 2\left(\frac{du_0}{dx} + \frac{d^3u_0}{dx^3}\right)\frac{\partial^2\phi}{\partial Y^2} - 2\left(\frac{d^2u_0}{dx^2} + \frac{d^4u_0}{dx^4}\right)\left(\frac{\partial\phi}{\partial Y}\right)^2, \tag{8.115}$$

where

$$\mathcal{L}_0 = \mu - \left(\frac{\partial^2}{\partial x^2} + 1\right)^2 - 3u_0^2. \tag{8.116}$$

Now

$$\mathcal{L}_0\frac{du_0}{dx} = \frac{\delta f}{\delta u}\frac{du_0}{dx} = 0, \tag{8.117}$$

from equation (8.105), and so $\mathcal{L}_0$ is a singular operator with du_0/dx in its kernel. $\mathcal{L}_0$ is also self-adjoint under periodic boundary conditions. Thus according to the Fredholm alternative theorem (see equation 7.20), equation (8.115) can only have a solution if

$$\frac{\partial\phi}{\partial T}\int_{-L/2}^{L/2}\left(\frac{du_0}{dx}\right)^2 dx = -\frac{\partial^2\phi}{\partial Y^2}\int_{-L/2}^{L/2} 2\frac{du_0}{dx}\left(\frac{du_0}{dx} + \frac{d^3u_0}{dx^3}\right)dx - \left(\frac{\partial\phi}{\partial Y}\right)^2\int_{-L/2}^{L/2} 2\frac{du_0}{dx}\left(\frac{d^2u_0}{dx^2} + \frac{d^4u_0}{dx^4}\right)dx, \tag{8.118}$$

where L is the spatial period of $u_0(x)$. The last integral vanishes, since

$$\int_{-L/2}^{L/2} 2\frac{du_0}{dx}\frac{d^2u_0}{dx^2}dx = \left[\left(\frac{du_0}{dx}\right)^2\right]_{-L/2}^{L/2} = 0, \tag{8.119}$$

$$\int_{-L/2}^{L/2} 2\frac{du_0}{dx}\frac{d^4u_0}{dx^4}dx = \left[2\frac{du_0}{dx}\frac{d^3u_0}{dx^3} - \left(\frac{d^2u_0}{dx^2}\right)^2\right]_{-L/2}^{L/2} = 0, \tag{8.120}$$

where the boundary terms vanish because $u_0(x)$ is periodic with period L. This leaves us with

$$\frac{\partial\phi}{\partial T} = -2\left(1 + \frac{\int_{-L/2}^{L/2}\frac{du_0}{dx}\frac{d^3u_0}{dx^3}dx}{\int_{-L/2}^{L/2}\left(\frac{du_0}{dx}\right)^2 dx}\right)\frac{\partial^2\phi}{\partial Y^2} \tag{8.121}$$

$$= -2\left(1 - \frac{\int_{-L/2}^{L/2}\left(\frac{d^2u_0}{dx^2}\right)^2 dx}{\int_{-L/2}^{L/2}\left(\frac{du_0}{dx}\right)^2 dx}\right)\frac{\partial^2\phi}{\partial Y^2}. \tag{8.122}$$

The zigzag instability sets in when

$$\frac{\int_{-L/2}^{L/2}\left(\frac{\mathrm{d}^2u_0}{\mathrm{d}x^2}\right)^2\mathrm{d}x}{\int_{-L/2}^{L/2}\left(\frac{\mathrm{d}u_0}{\mathrm{d}x}\right)^2\mathrm{d}x}<1 \tag{8.123}$$

For a roll solution of form $u_o(x)=R\cos(1+\delta q)x$, where R and q are constant and $0<\delta\ll 1$, this instability criterion reduces to

$$(1+\delta q)^2<1, \tag{8.124}$$

$$q+O(\delta)<0. \tag{8.125}$$

This agrees with the zigzag instability boundary found in Section 8.3 for rolls with wavenumbers close to critical, even though the rolls here are fully nonlinear, whereas in Section 8.3 they have small amplitude since the analysis is carried out close to onset of the initial pattern-forming instability.

In the zigzag unstable region equation (8.122) is ill-posed, so higher order terms are needed to regularise it. To do this we work close to the instability boundary, setting

$$\frac{\int_{-L/2}^{L/2}\left(\frac{\mathrm{d}^2u_0}{\mathrm{d}x^2}\right)^2\mathrm{d}x}{\int_{-L/2}^{L/2}\left(\frac{\mathrm{d}u_0}{\mathrm{d}x}\right)^2\mathrm{d}x}=1+\epsilon^2\frac{D}{2}, \tag{8.126}$$

where D is a constant that is negative in the zigzag-unstable regime. We also rescale $T=\epsilon^4 t$ to ensure that the leading order phase equation contains higher order linear and nonlinear terms. Again if the appropriate scalings were not known in advance it would be possible to work them out through trial and error.

At $O(1)$ in the rescaled expansion we find, as before, that equation (8.114) must be satisfied. Now the $O(\epsilon^2)$ terms give

$$\mathcal{L}_0u_1=2\left(\frac{\mathrm{d}u_0}{\mathrm{d}x}+\frac{\mathrm{d}^3u_0}{\mathrm{d}x^3}\right)\frac{\partial^2\phi}{\partial Y^2}+2\left(\frac{\mathrm{d}^2u_0}{\mathrm{d}x^2}+\frac{\mathrm{d}^4u_0}{\mathrm{d}x^4}\right)\left(\frac{\partial\phi}{\partial Y}\right)^2, \tag{8.127}$$

which according to the Fredholm alternative theorem will have a solution if

$$2\int_{-L/2}^{L/2}\frac{\mathrm{d}u_0}{\mathrm{d}x}\left(\frac{\mathrm{d}u_0}{\mathrm{d}x}+\frac{\mathrm{d}^3u_0}{\mathrm{d}x^3}\right)\frac{\partial^2\phi}{\partial Y^2}+\frac{\mathrm{d}u_0}{\mathrm{d}x}\left(\frac{\mathrm{d}^2u_0}{\mathrm{d}x^2}+\frac{\mathrm{d}^4u_0}{\mathrm{d}x^4}\right)\left(\frac{\partial\phi}{\partial Y}\right)^2\mathrm{d}x=0 \tag{8.128}$$

at $O(1)$. Using equations (8.119) and (8.120) and integrating by parts, this reduces to the requirement that

$$2\left(\int_{-L/2}^{L/2}\left(\frac{\mathrm{d}u_0}{\mathrm{d}x}\right)^2\mathrm{d}x-\int_{-L/2}^{L/2}\left(\frac{\mathrm{d}^2u_0}{\mathrm{d}x^2}\right)^2\mathrm{d}x\right)\frac{\partial^2\phi}{\partial Y^2}=0 \tag{8.129}$$

at $O(1)$. By equation (8.126) we have

$$2\left(\int_{-L/2}^{L/2}\left(\frac{\mathrm{d}u_0}{\mathrm{d}x}\right)^2\mathrm{d}x-\int_{-L/2}^{L/2}\left(\frac{\mathrm{d}^2u_0}{\mathrm{d}x^2}\right)^2\mathrm{d}x\right)\frac{\partial^2\phi}{\partial Y^2}=\epsilon^2D\int_{-L/2}^{L/2}\left(\frac{\mathrm{d}^2u_0}{\mathrm{d}x^2}\right)^2\mathrm{d}x \tag{8.130}$$

and so the solvability condition is satisfied. However there will be a contribution from the righthand side of equation (8.127) to the solvability condition at the next order in the expansion. We can write the solution of equation (8.127) in the form

$$u_1(x,Y,T)=A(x)\frac{\partial^2\phi}{\partial Y^2}+B(x)\left(\frac{\partial\phi}{\partial Y}\right)^2, \tag{8.131}$$

where $A(x)$ and $B(x)$ satisfy

$$\mathcal{L}_0A=2\left(\frac{\mathrm{d}u_0}{\mathrm{d}x}+\frac{\mathrm{d}^3u_0}{\mathrm{d}x^3}\right), \tag{8.132}$$

$$\mathcal{L}_0B=2\left(\frac{\mathrm{d}^2u_0}{\mathrm{d}x^2}+\frac{\mathrm{d}^4u_0}{\mathrm{d}x^4}\right) \tag{8.133}$$

Since $u_0(x)$ is a roll solution it has reflection symmetry about some axis, which we shall take to be $x=0$, without loss of generality. Consequently $u_0(x)$ and all its even x derivatives are even in x, while all its odd x derivatives are odd in x. Since $\mathcal{L}_0$ is even in x, we can deduce from equations (8.132) and (8.133) that $A(x)$ is odd and $B(x)$ is even. These observations are useful at the next stage of the expansion.

At $O(\epsilon^4)$ we find

$$\begin{aligned}\frac{\mathrm{d}u_0}{\mathrm{d}x}\frac{\partial\phi}{\partial T}=\mathcal{L}_0u_2-3u_0u_1^2-2\epsilon^{-2}\left(\frac{\mathrm{d}u_0}{\mathrm{d}x}+\frac{\mathrm{d}^3u_0}{\mathrm{d}x^3}\right)\frac{\partial^2\phi}{\partial Y^2}\\ -2\left(1+\frac{\partial^2}{\partial x^2}\right)\frac{\partial^2u_1}{\partial Y^2}-\frac{\mathrm{d}^4u_0}{\mathrm{d}x^4}\left(\frac{\partial\phi}{\partial Y}\right)^4\\ -6\frac{\mathrm{d}^3u_0}{\mathrm{d}x^3}\left(\frac{\partial\phi}{\partial Y}\right)^2\frac{\partial^2\phi}{\partial Y^2}-\frac{\mathrm{d}^2u_0}{\mathrm{d}x^2}\left(3\left(\frac{\partial^2\phi}{\partial Y^2}\right)^2+4\frac{\partial\phi}{\partial Y}\frac{\partial^3\phi}{\partial Y^3}\right)\\ -\frac{\mathrm{d}u_0}{\mathrm{d}x}\frac{\partial^4\phi}{\partial Y^4},\end{aligned} \tag{8.134}$$

where we retain the term with the factor of ϵ^{-2} to account for the $O(\epsilon^2)$ contribution from equation (8 130). After some calculation, the Fredholm solvability condition gives

$$\begin{aligned}\frac{\partial\phi}{\partial T}\int_{-L/2}^{L/2}\left(\frac{\mathrm{d}u_0}{\mathrm{d}x}\right)^2\mathrm{d}x = &\int_{-L/2}^{L/2}3\frac{\mathrm{d}u_0}{\mathrm{d}x}u_0u_1^2\mathrm{d}x\\ &+\left(D\frac{\partial^2\phi}{\partial Y^2}+\frac{\partial^4\phi}{\partial Y^4}\right)\int_{-L/2}^{L/2}\left(\frac{\mathrm{d}u_0}{\mathrm{d}x}\right)^2\mathrm{d}x\\ &+2\int_{-L/2}^{L/2}\frac{\mathrm{d}u_0}{\mathrm{d}x}\left(1+\frac{\partial^2}{\partial x^2}\right)\frac{\partial^2u_1}{\partial Y^2}\mathrm{d}x\\ &-6\left(\frac{\partial\phi}{\partial Y}\right)^2\frac{\partial^2\phi}{\partial Y^2}\int_{-L/2}^{L/2}\left(\frac{\mathrm{d}^2u_0}{\mathrm{d}x^2}\right)^2\mathrm{d}x.\end{aligned} \tag{8.135}$$

We now substitute for u_1 from equation (8.131) and evaluate all the integrals. The following integrals vanish because their integrands are odd,

$$\int_{-L/2}^{L/2}\frac{\mathrm{d}u_0}{\mathrm{d}x}\left(1+\frac{\partial^2}{\partial x^2}\right)B(x)\mathrm{d}x=0, \tag{8.136}$$

$$\int_{-L/2}^{L/2}u_0\frac{\mathrm{d}u_0}{\mathrm{d}x}A(x)^2\mathrm{d}x=0, \tag{8.137}$$

$$\int_{-L/2}^{L/2}u_0\frac{\mathrm{d}u_0}{\mathrm{d}x}B(x)^2\mathrm{d}x=0, \tag{8.138}$$

and the phase equation reduces to

$$\frac{\partial\phi}{\partial T}=D\frac{\partial^2\phi}{\partial Y^2}-\kappa\frac{\partial^4\phi}{\partial Y^4}+g\left(\frac{\partial\phi}{\partial Y}\right)^2\frac{\partial^2\phi}{\partial Y^2}, \tag{8.139}$$

where κ and g are real constants that can be evaluated in terms of combinations of the integrals in equation (8 135). In applications κ is typically positive, while g is also usually positive in convection problems The phase equation takes the same form as equation (8 80), derived for small-amplitude solutions using the Newell–Whitehead–Segel equation, since the terms that are permitted to appear are determined by symmetry. However the coefficients may be different.

A similar method can be used to calculate the Eckhaus instability boundary for finite amplitude rolls Writing

$$u(x,t)=u_0(x+\phi(X,T))+\epsilon u_1(x,X,T)+\epsilon^2u_2(x,X,T)+\cdots, \tag{8.140}$$

where $X = \epsilon x$ and $T = \epsilon^2 t$, substituting into equation (8.110) and equating terms at each order in ϵ gives the leading order phase-diffusion equation

$$\frac{\partial \phi}{\partial T} = D \frac{\partial^2 \phi}{\partial X^2}, \tag{8.141}$$

where

$$D = \frac{\int_{-L/2}^{L/2} 6\left(\frac{\mathrm{d}^2 u_0}{\mathrm{d}x^2}\right)^2 - 2\left(\frac{\mathrm{d}u_0}{\mathrm{d}x}\right)^2 - 4\frac{\mathrm{d}u_0}{\mathrm{d}x}\left(\frac{\mathrm{d}^3 F}{\mathrm{d}x^3} + \frac{\mathrm{d}F}{\mathrm{d}x}\right)\mathrm{d}x}{\int_{-L/2}^{L/2}\left(\frac{\mathrm{d}u_0}{\mathrm{d}x}\right)^2 \mathrm{d}x}, \tag{8.142}$$

$$\mathcal{L}_0 F = 4\left(\frac{\mathrm{d}^2 u_0}{\mathrm{d}x^2} + \frac{\mathrm{d}^4 u_0}{\mathrm{d}x^4}\right) \tag{8.143}$$

The Eckhaus instability sets in for $D < 0$.

8.5 The cross-roll instability

In contrast to the Eckhaus and zigzag instabilities, which are associated with the phase of the rolls, the cross-roll instability (Busse, 1962; Schlüter, Lortz & Busse, 1965; Newell & Whitehead, 1969; Busse & Whitehead, 1971) is an amplitude instability. It consists of a set of rolls growing at an angle to the original pattern, as shown in Figure 8.9. Strictly the instability is only called cross-roll if the new set of rolls is at right angles to the original pattern. Otherwise, it is called an oblique-roll instability. In this case we can write the leading order solution in the form

$$u(x, y, t) = A(X, Y, T)e^{ix} + B(\widetilde{X}, \widetilde{Y}, \widetilde{T})e^{i\widetilde{x}} + c.c\,, \tag{8.144}$$

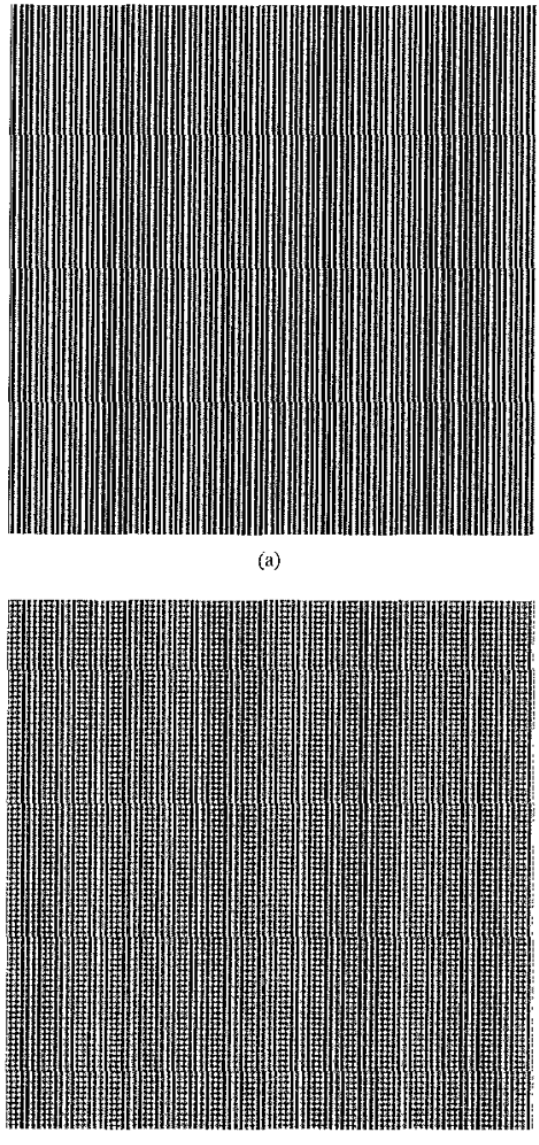
where the $\widetilde{x}$ direction is at some angle, θ, to the x direction and the two variables have the same scaling. Here and below all the tilde variables bear the same relation to $\widetilde{x}$ as the plain variables do to x. In a Euclidean-symmetric system we have

$$\frac{\partial A}{\partial T} = \mu A - |A|^2 A - \alpha |B|^2 A + \frac{\partial^2 A}{\partial X^2}, \tag{8.145}$$

$$\frac{\partial B}{\partial T} = \mu B - |B|^2 B - \alpha |A|^2 B + \frac{\partial^2 B}{\partial \widetilde{X}^2}. \tag{8.146}$$

The coupling α between the two sets of rolls will be dependent both upon the angle between them and upon the details of the particular system under consideration.

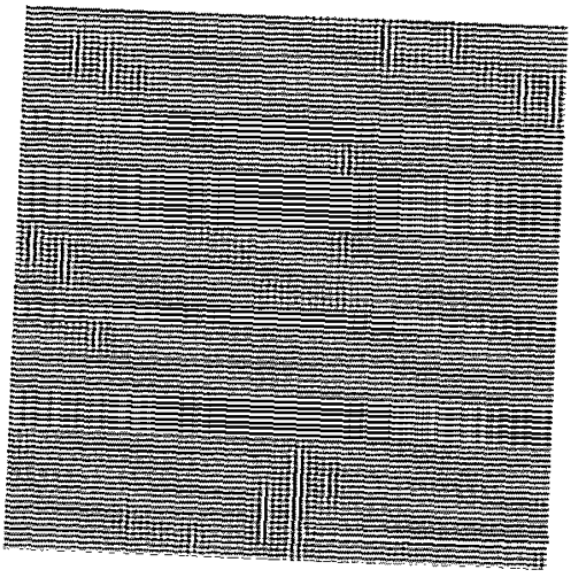
Initially, we have a set of rolls, $A = R_0 e^{iqX}$, where $R_0^2 = \mu - q^2$. We allow a small cross-roll perturbation, so that $B = b(T)e^{ip\widetilde{X}}$, where $|b| \ll R_0$. Linearising



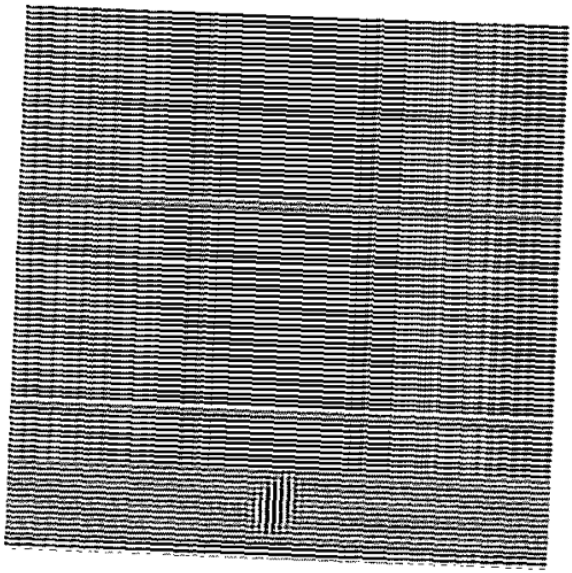
(a)

(b)

Fig. 8.9 The cross-roll instability in equations (8.145) and (8.146), with periodic boundary conditions and an angle of $\pi/2$ between the two sets of rolls. The pictures show greyscale plots of $u(x, y, t) = A(X, Y, T)e^{ix} + B(X, Y, T)e^{iy} + c.c.$, where $\epsilon \equiv X/x = Y/y = 0.20$. (a) The initial state at time $T = 0$ is a uniform roll solution with $q = 1.0$ and $\mu = 3.5$ in the cross-roll unstable region where the pattern wavelength is too short. A small random perturbation has been added to the original rolls, and a small random real cross-roll perturbation, B, has also been added. (b) At $T = 4$ cross-rolls have started to grow. (c) By $T = 16$, the original rolls are dying away in patches. (d) Finally at $T = 39$, the cross-rolls have invaded almost the entire domain. Though it is hard to see from the pictures, the final wavelength is in fact longer than the original.



(c)



(d)

Fig 8 9 (*cont*)

in b, we find that

$$\frac{\mathrm{d}b}{\mathrm{d}T} = (\mu - p^2 - \alpha(\mu - q^2))b. \tag{8 147}$$

For a fixed q, the fastest growing mode has $p = 0$, and it is unstable when $\mu(1 - \alpha) > -\alpha q^2$

If $\alpha < 1$, then there is instability for

$$\mu > \mu_c \equiv \frac{-\alpha q^2}{1-\alpha} \tag{8.148}$$

Now $\mu_c < q^2$, so rolls are always unstable to cross-rolls. However, in this case rolls are not the preferred solution at the primary bifurcation from the conduction solution. When the angle between the two sets of rolls is $\pi/2$, as seen in convection (Busse & Whitehead 1971), and the system has square symmetry, the amplitude equations (8.145) and (8.146) show that squares $|A|^2 = |B|^2$ are the stable, preferred solution. If the angle is not $\pi/2$, then again the pattern $|A|^2 = |B|^2$ is preferred, but these are now rectangles.

If $\alpha > 1$, then rolls are unstable to cross-rolls for

$$\mu < \mu_c \equiv \frac{-\alpha q^2}{1-\alpha} \tag{8.149}$$

In this case $\mu_c > q^2$, and so for any μ there is a band of stable wavenumbers around $q = 0$. As for the Eckhaus mode, instability occurs when the roll wavelength is too short or too long. The instability boundary is parabolic in (μ, q) space. Its position in relation to the Eckhaus instability boundary depends upon the value of α; if $\alpha < 3/2$, the cross-roll instability comes in at a smaller value of q^2 than the Eckhaus instability, whereas if $\alpha > 3/2$, the Eckhaus instability comes in first. The cross-roll unstable region looks like the Eckhaus unstable region in Figure 8.1a, except that now the upper stability boundary is $\mu = -\alpha q^2/(1-\alpha)$.

In Rayleigh–Bénard convection at high Prandtl number, the cross-roll instability comes in first, while for low Prandtl number fluids the Eckhaus instability is the more dangerous (see the discussion of the Busse stability balloon at the end of the next section). The calculation of the cross-roll instability boundary when the governing equation is the Swift–Hohenberg equation is set as an exercise at the end of the chapter.

8.6 Prandtl-number-dependent instabilities of convection rolls

As well as the universal instabilities – Eckhaus, zigzag and cross-roll – that can be understood in the context of a single vertical velocity or density perturbation,

$$u_z(x, y, t) = \sum_j A_j(X, Y, T) e^{ik_j x} + c.c. + h.o.t., \tag{8.150}$$

with amplitudes A_j evolving according to equations determined by symmetry, convection rolls in systems with appropriate fluid properties can also be subject to instabilities that arise from the presence of a mean fluid flow. This drift flow

is damped for no-slip boundary conditions ($\boldsymbol{u} = \mathbf{0}$ on the top and bottom plates of the convection cell, where $\boldsymbol{u}$ is the fluid velocity), with stronger damping for higher Prandtl numbers. With stress-free boundaries ($u_z = \partial^2 u_z/\partial z^2 = 0$ on the top and bottom plates, where u_z is the vertical component of the fluid velocity and z is the vertical coordinate) a uniform drift flow is neutral, since the Boussinesq equations and the boundary conditions are Galilean-invariant, in other words unchanged under the transformation $\boldsymbol{u} \to \boldsymbol{u} + \boldsymbol{v}_h$, where $\boldsymbol{v}_h$ is a constant horizontal velocity. Stress-free boundaries can't be achieved in real experiments, but for low Prandtl number fluids, such as liquid metals, the drift flow is only weakly damped (owing to the development of viscous boundary layers close to the plates of the convection cell), and mean flows are observed for Prandtl numbers smaller than $O(1)$. In the limit of infinite Prandtl number, the generation of vertical vorticity, corresponding to a nonuniform mean drift flow, is suppressed even in the stress-free case. We consider a neutral mean drift mode to be present for low Prandtl number fluids with experimentally realisable boundaries. The coupling of the horizontal mean drift to the temperature and vertical velocity perturbations gives rise to the **skew-varicose** and **oscillatory** instabilities of rolls.

Horizontal drift flows are associated with nonzero vertical vorticity. A straight roll solution

$$u_z(x, y, t) = Ae^{ix} + c.c., \tag{8.151}$$

where A is a constant, does not generate vertical vorticity, as can be seen from the vertical component of the vorticity equation (1.17), so the mean drift flow can only arise from long-scale modulations of A, when $A \equiv A(X, Y, T)$, and must therefore be a large-scale ($k = 0$) mode itself. Since there is a neutral $k = 0$ mode present, the Ginzburg–Landau and Newell–Whitehead–Segel equations do not hold.

As stated in Chapter 1, an incompressible fluid flow with velocity $\boldsymbol{U}(x, y, t)$ satisfies

$$\nabla \cdot \boldsymbol{U} = 0 \tag{8.152}$$

A two-dimensional incompressible flow such as the mean drift can therefore be written

$$\boldsymbol{U} = \nabla \times \hat{z}\psi, \tag{8.153}$$

where $\psi(x, y, t)$ is a so-called **streamfunction** associated with the mean drift flow, and $\hat{z}$ is a unit vector in the z direction. At linear order, the vorticity, $\omega = \nabla \times \boldsymbol{U}$, satisfies the equation

$$\frac{\partial \omega}{\partial t} = \nu \nabla^2 \omega, \tag{8.154}$$

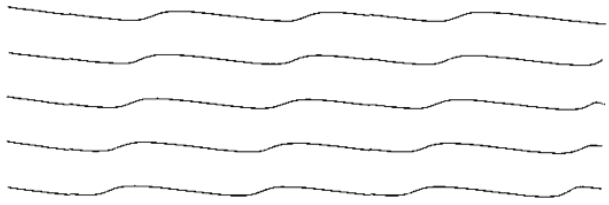


Fig 8.10 Rolls undergoing the skew-varicose instability

where ν is the kinematic viscosity of the fluid For a two-dimensional incompressible flow, we have

$$\omega = -\hat{z}\nabla^2\psi, \tag{8.155}$$

and so the vorticity equation can be written

$$\frac{\partial}{\partial t}\nabla^2\psi = \nu\nabla^4\psi. \tag{8.156}$$

This equation will be coupled to the evolution equation for the roll amplitude, $A(X, Y, T)$ The nonlinear coupling terms will involve the derivatives of ψ, corresponding to the mean drift flow and its derivatives, along with the amplitude, A, and its derivatives

8.6.1 The skew-varicose instability

The skew-varicose instability looks like an Eckhaus instability that takes place at an angle to the original roll axes, as shown in Figure 8.10; for experimental observations, see Busse and Clever (1979) The analysis below reveals that there are both stationary and oscillatory skew-varicose modes. The relevant symmetries are

(i) translation in x: $\{x \to x + x_0, A \to Ae^{-ix_0}, x_0 \text{ constant}\}$;
(ii) reflection in x: $\{X \to -X, A \to \bar{A}, \psi \to -\psi\}$;
(iii) reflection in y: $\{Y \to -Y, \psi \to -\psi\}$;
(iv) translation in X: $\{X \to X + X_0, A(X, Y, T) \to A(X - X_0, Y, T), \psi(X, Y, T) \to \psi(X - X_0, Y, T), X_0 \text{ constant}\}$;
(v) translation in Y: $\{Y \to Y + Y_0, A(X, Y, T) \to A(X, Y - Y_0, T), \psi(X, Y, T) \to \psi(X, Y - Y_0, T), Y_0 \text{ constant}\}$;
(vi) translation in T: $\{T \to T + T_0, A(X, Y, T) \to A(X, Y, T - T_0), \psi(X, Y, T) \to \psi(X, Y, T - T_0), T_0 \text{ constant}\}$

The system must also be insensitive to changing the streamfunction by a constant value, as the mean drift depends only on its derivatives, so we must have equivariance under $\psi \to \psi + \psi_0$, where ψ_0 is a constant

Attempts to describe the skew-varicose instabilities using amplitude equations are not completely satisfactory. It appears that two different calculations with different scalings for X and Y are needed to capture all the behaviour. Zippelius and Siggia (1983) used scalings

$$X = \epsilon x, \quad Y = \epsilon^{\frac{1}{2}} y \tag{8.157}$$

just as for the analysis of the Eckhaus and zigzag instabilities. The self-consistent scalings

$$A \sim \epsilon, \quad \psi \sim \epsilon^{\frac{3}{2}}, \quad \mu \sim \epsilon^2, \quad T = \epsilon^2 t, \tag{8.158}$$

lead to the governing equations

$$\frac{\partial A}{\partial T} = \mu A - |A|^2 A + \left(\frac{\partial}{\partial X} - \frac{i}{2}\frac{\partial^2}{\partial Y^2}\right)^2 A - iA\frac{\partial \psi}{\partial Y}, \tag{8.159}$$

$$0 = \nu \frac{\partial^2 \omega}{\partial Y^2} + \alpha \frac{\partial}{\partial Y}\left\{\bar{A}\left(\frac{\partial}{\partial X} - \frac{i}{2}\frac{\partial^2}{\partial Y^2}\right)A + c.c.,\right\} \tag{8.160}$$

$$\omega = -\frac{\partial^2 \psi}{\partial Y^2}, \tag{8.161}$$

where $\omega = |\boldsymbol{\omega}|$. Zippelius and Siggia (1983) derived equivalent equations, except that they retained small X-derivative terms in $\nabla^2\omega$. Equations (8.159)–(8.161) have a further symmetry in addition to those mentioned above: they are equivariant under transformation to a frame moving with constant velocity,

$$x \to x - vT, \quad A \to Ae^{-ivT}, \quad \psi \to \psi + vY, \tag{8.162}$$

where v is a real constant. This **Galilean** symmetry will be discussed in more detail in Section 9.3. Its presence here reflects the invariance of the governing Oberbeck–Boussinesq equations and stress-free boundary conditions under transformation to a frame moving in any horizontal direction with constant speed, as mentioned above.

For disturbances with no Y dependence, the amplitude equation (8.159) decouples from the vorticity equation (8.161), so the Eckhaus instability is not affected by the presence of the mean flow. It remains to investigate the zigzag instability (where there is no X dependence in the perturbation) and the skew-varicose instabilities where there is variation in both directions.

Stationary roll solutions, $A = R_0 e^{iqX}$ with $R_0^2 = \mu - q^2$ give rise to no large scale flow, so $\psi = 0$. Perturbing around this solution, such that $A = R_0(1 + r)$

$e^{i(qX+\theta)}$, where $|r|, |\theta|, |\psi| \ll 1$ gives

$$\frac{\partial r}{\partial T} = -2R_0^2 r + \frac{\partial^2 r}{\partial X^2} + q\frac{\partial^2 r}{\partial Y^2} - \frac{1}{4}\frac{\partial^4 r}{\partial Y^4} - 2q\frac{\partial \theta}{\partial X} + \frac{\partial^3 \theta}{\partial X \partial Y^2}, \tag{8.163}$$

$$\frac{\partial \theta}{\partial T} = 2q\frac{\partial r}{\partial X} - \frac{\partial^3 r}{\partial X \partial Y^2} + \frac{\partial^2 \theta}{\partial X^2} + q\frac{\partial^2 \theta}{\partial Y^2} - \frac{1}{4}\frac{\partial^4 \theta}{\partial Y^4} - \frac{\partial \psi}{\partial Y}, \tag{8.164}$$

$$0 = -\nu\frac{\partial^4 \psi}{\partial Y^4} + 2\alpha R_0^2\left(\frac{\partial^2 r}{\partial X \partial Y} + \frac{1}{2}\frac{\partial^3 \theta}{\partial Y^3}\right). \tag{8.165}$$

Now setting $r = \widehat{r}e^{\sigma T + ikX + ilY}$, $\theta = \widehat{\theta}e^{\sigma T + ikX + ilY}$ and $\psi = \widehat{\psi}e^{\sigma T + ikX + ilY}$, where $\widehat{r}$, $\widehat{\theta}$ and $\widehat{\psi}$ are real constants leads to a dispersion relation

$$\begin{aligned}
&\sigma^2 + \sigma\left(2R_0^2 + 2k^2 + 2ql^2 + \frac{1}{2}l^4 + \frac{\alpha R_0^2}{\nu}\right) \\
&+ \left(2R_0^2 + k^2 + ql^2 + \frac{1}{4}l^4\right)\left(\frac{\alpha}{\nu}R_0^2 + k^2 + ql^2 + \frac{1}{4}l^4\right) \\
&- k^2(2q + l^2)\left(2q + l^2 + \frac{2\alpha R_0^2}{\nu l^2}\right) = 0.
\end{aligned} \tag{8.166}$$

We are assuming that the perturbation has some Y-dependence, so $l \neq 0$ and the final term is not singular. We have to take the long-wave limit carefully here, because there is a term proportional to k^2/l^2 in the dispersion relation, and its behaviour in the limit determines the order at which it contributes to the growth rate eigenvalues. If we take $k \sim l^2 \to 0$, appropriate to the relative scalings of X and Y in equation (8.157), then the growth rate eigenvalues are

$$\sigma_1 = -2R_0^2, \tag{8.167}$$

$$\sigma_2 = -\frac{\alpha R_0^2}{\nu}, \tag{8.168}$$

at leading order. The first eigenvalue, σ_1, is associated with the decay of amplitude perturbations, while σ_2 shows that the phase modes are stable at leading order for $\alpha > 0$ (as is the case for Rayleigh–Bénard convection – see Zippelius & Siggia, 1983). In particular, the zigzag instability is strongly damped for small Prandtl number (corresponding here to small ν). Physically, this is because the large scale flow is generated in the opposite direction to the bends of the zigzags, and so tends to push them back in, as shown in Figure 8.11. To see why this is so, consider Fourier mode phase and streamfunction perturbations as above. According to

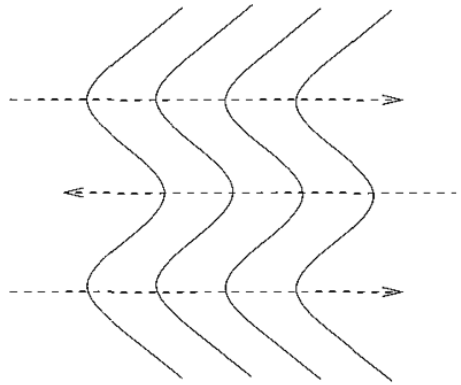


Fig 8 11 The zigzag instability is suppressed by the large scale flow generated by the bends in the rolls The direction of the mean flow is shown by the dashed lines

equation (8.165) the mean flow $\boldsymbol{U} = (U_x, U_y)$ satisfies

$$U_x = \frac{\partial \psi}{\partial Y} = \frac{\alpha R_0^2}{\mu}\theta. \tag{8.169}$$

If the phase, θ, is positive the rolls are shifted left, but the velocity U_x, in the x-direction, is also positive and pushes the rolls to the right. Conversely if the phase shift is to the right ($\theta < 0$) the mean flow pushes left. So the mean flow resists bends in the rolls

In contrast if $k \sim l \to 0$ then at leading order the dispersion relation becomes

$$\sigma^2 + \sigma R_0^2\left(2 + \frac{\alpha}{\nu}\right) + 2R_0^2\frac{\alpha}{\nu}\left(R_0^2 - 2q\frac{k^2}{l^2}\right) = 0, \tag{8.170}$$

so there is instability if

$$\alpha\left(R_0^2 - 2q\frac{k^2}{l^2}\right) < 0 \tag{8.171}$$

For $\alpha > 0$ all rolls with $q > 0$ will be unstable to modes with large enough k^2/l^2 For $\alpha < 0$, all rolls with $q < 0$ will be unstable for all k^2/l^2, and rolls with $q > 0$ will be unstable to modes with k^2/l^2 small enough.

The fact that the predicted behaviour in the $\alpha > 0$ case changes from stability to instability as the ratio k^2/l^2 increases, and that $\sigma^2 \to \infty$ as $l \to 0$, suggests that the skew-varicose instability is not fully captured using the scalings (8.157) and (8.158).

An alternative scaling that is appropriate when $k \sim l$ is

$$X = \epsilon x, Y = \epsilon y \tag{8.172}$$

We can add the standard scalings

$$A \sim \epsilon, \quad \mu \sim \epsilon^2, \quad I = \epsilon^2 t, \tag{8.173}$$

but there is now no consistent scaling for ψ The leading order governing equations are

$$\frac{\partial A}{\partial T} = \mu A - |A|^2 A + \frac{\partial^2 A}{\partial X^2} - iA\frac{\partial \psi}{\partial Y}, \tag{8.174}$$

$$\frac{\partial \omega}{\partial T} = \nu \nabla^2 \omega + \alpha \frac{\partial^2 |A|^2}{\partial X \partial Y}, \tag{8.175}$$

$$\omega = -\nabla^2 \psi \tag{8.176}$$

The amplitude equation suggests the scaling $\psi \sim \epsilon$, but this means that the coupling term in the vorticity equation comes in at $O(\epsilon^4)$, while the remaining terms come in at $O(\epsilon^5)$ These scalings were none the less adopted by Busse and Bolton (1984) and Bernoff (1994) Bernoff includes higher order terms in the governing equations in an attempt to reconcile the Zippelius and Siggia (1983) and Busse and Bolton (1984) results

If we analyse the stability of rolls using the new governing equations we find the dispersion relation

$$\begin{aligned}
&\sigma^3 + \sigma^2(2R_0^2 + \nu(k^2 + l^2) + 2k^2) \\
&+\sigma(2\nu(R_0^2 + k^2)(k^2 + l^2) + 2k^2(R_0^2 - 2q^2) + k^4) \\
&+\nu k^4(k^2 + l^2) + 2\nu(R_0^2 - 2q^2)k^2(k^2 + l^2) - 4\alpha q R_0^2 \frac{k^2 l^2}{k^2 + l^2} = 0
\end{aligned} \tag{8.177}$$

In the long-wave limit $k \sim l \to 0$ there are three roots

$$\sigma_1 = -2R_0^2 + O(k^2), \tag{8.178}$$

$$\sigma_{2,3} = \pm\sqrt{2\alpha q}\frac{kl}{(k^2 + l^2)^{\frac{1}{2}}} + k^2\left(-\frac{\alpha q}{2R_0^2}S - \frac{\nu}{2(1-S)} - \frac{1}{2} + \frac{q^2}{R_0^2}\right) + O(k^3), \tag{8.179}$$

where $S = l^2/(k^2 + l^2)$ and hence $0 \le S \le 1$ The first eigenvalue, as always, is associated with the rapid relaxation of the amplitude mode to its equilibrium value The remaining two phase mode eigenvalues are either real, one positive and one negative at $O(k)$, if $\alpha q > 0$, or a complex conjugate pair if $\alpha q < 0$, with growth rate determined at $O(k^2)$ If $\alpha q > 0$, the rolls are therefore unstable to the monotone skew-varicose instability If $-\nu R_0^2 < \alpha q < 0$ the $O(k^2)$ growth

rate has a maximum

$$\sigma_* = k^2 \left\{ \frac{q^2}{R_0^2} - \frac{1}{2}(1+\nu) \right\} \tag{8.180}$$

at $S = 0$, and so will be positive if

$$q^2 > \frac{1}{2}(\nu+1)R_0^2. \tag{8.181}$$

If $\alpha q < -\nu R_0^2$ the maximum growth rate is

$$\sigma_* = -\frac{k^2}{4R_0^2} \left\{ 2\alpha q \left(1 - 2\sqrt{\frac{-\nu R_0^2}{\alpha q}} \right) + 2(R_0^2 - 2q^2) \right\}, \tag{8.182}$$

at $(1-S)^2 = -\nu R_0^2/\alpha q$, and is positive for sufficiently large $|q|$. In either $\alpha q < 0$ case, if the growth rate is positive, there is an oscillatory skew-varicose instability

A wide range of different scalings has been used to investigate the skew-varicose instability, though there is still no real agreement on which is most appropriate. For example, Bernoff (1994) set $q \sim k^2 \sim l^2$ to get monotonic and oscillatory skew-varicose stability boundaries of the form $\mu \propto q$. Mielke (1997) presented an analysis specific to Rayleigh–Bénard convection, and concluded that combining results from both scalings was necessary to describe the skew-varicose instability fully.

8.6.2 *The oscillatory instability of rolls*

The oscillatory instability takes the form of transverse wiggles or zigzags that propagate along the roll axes, as shown in Figure 8.12; for experimental observations, see Krishnamurti (1970) and Willis and Deardorff (1970) and for early theoretical work Busse (1972). The oscillatory instability can be understood in the context of a system with Galilean, translation and reflection symmetries. The linear analysis is simpler than that for the skew-varicose instability since the rolls are modulated only in the Y direction, so there are no complications over the relative scalings of X and Y.

For steady rolls, we have a solution $u_0(x)$ for the fluid velocity perpendicular to the roll axes. A perturbation of this solution in the phase and velocity gives

$$u(x, y, t) = u_0(x + \phi(Y, T)) + V(Y, T), \tag{8.183}$$

where the phase perturbation, ϕ, and the velocity perturbation, V, depend on long scales Y and T in agreement with what is observed in the oscillatory instability.

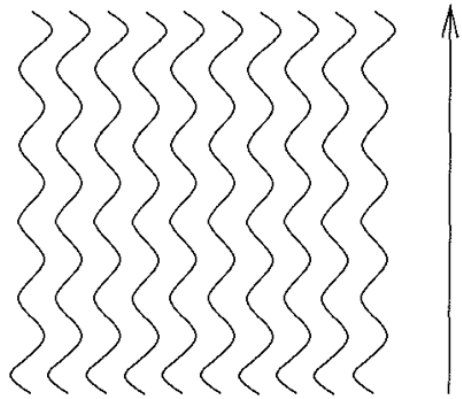


Fig 8 12 Rolls undergoing the oscillatory instability. The arrow indicates the direction of travel of the wiggles

This and the following analysis is a simplification of the work of Fauve, Boulton and Brachet (1987) We can derive evolution equations for ϕ and V using equivariance under the symmetries

(i) translation in x: $\{x \to x + x_0, \phi \to \phi - x_0, x_0 \text{ constant}\}$;
(ii) transformation to a moving frame: $\{x \to x + vT, \phi \to \phi - vT, V \to V + v, v$ constant$\}$;
(iii) reflection in x: $\{x \to -x, \phi \to -\phi, V \to -V\}$;
(iv) reflection in y: $\{Y \to -Y\}$;
(v) translation in Y: $\{Y \to Y + Y_0, \phi(Y, T) \to \phi(Y - Y_0, T), V(Y, T) \to V(Y - Y_0, T), Y_0 \text{ constant}\}$;
(vi) translation in T: $\{T \to T + T_0, \phi(Y, T) \to \phi(Y, T - T_0), V(Y, T) \to V(Y, T - T_0), T_0 \text{ constant}\}$.

We can deduce the amplitude equations

$$\frac{\partial \phi}{\partial T} = -V + a\frac{\partial^2 \phi}{\partial Y^2} + \cdots , \tag{8 184}$$

$$\frac{\partial V}{\partial T} = -b\frac{\partial^2 \phi}{\partial Y^2} + c\frac{\partial^2 V}{\partial Y^2} + \cdots , \tag{8 185}$$

to the first two orders in ϵ, where we scale $T = \epsilon t$, $Y = \epsilon y$, $V/\phi \sim \epsilon$ and where a, b and c are real $O(1)$ constants Setting $\phi = \hat{\phi} e^{\sigma T + ikY}$ and $V = \hat{V} e^{\sigma T + ikY}$, where $\hat{\phi}$ and $\hat{V}$ are constants, leads in the long-wave limit ($k \to 0$) to the growth rate eigenvalues

$$\sigma = \pm i\sqrt{b}k - \frac{(a + c)}{2}k^2 + O(k^3). \tag{8 186}$$

If $a + c$ is negative and b is positive then the roll solution is unstable to an oscillatory mode. If b is negative on the other hand, then there is a growing steady zigzag mode at small $|k|$. Busse (1972) showed by direct calculation from the governing equations that conditions for an oscillatory instability are satisfied in Rayleigh–Bénard convection. Fauve, Bolton and Brachet (1987) make a phenomenological extension of the analysis to the case of no-slip boundary conditions, and show that this implies a finite wavenumber and frequency at onset. Whether the oscillatory mode corresponds to a propagating wave or a standing wave is only determined at nonlinear order: Fauve , Bolton and Brachet show that a propagating wave is selected, corresponding to bends that travel along the roll axis; an exercise at the end of this chapter takes you through some of their analysis.

8.6.3 The Busse balloon

The region in (R, k, σ) space corresponding to stable rolls in Boussinesq Rayleigh–Bénard convection is known as the **Busse balloon**, where here R is the Rayleigh number and σ is the Prandtl number, both introduced in Chapter 1, while $k = k_c + \epsilon q$ is the roll wavenumber. Its boundaries were established by Busse and his coworkers in a series of calculations and experiments published over many years (see, for example, Busse and Whitehead, 1971). The bifurcation parameter, μ, is related to the Rayleigh number by

$$\mu = \frac{R - R_c}{\epsilon^2 R_c} \tag{8.187}$$

where R_c is the critical Rayleigh number for the onset of convection, so a section through the Busse balloon at fixed Prandtl number defines the region of stable rolls in (q, μ) space. By analogy the region of stable stripes in (q, μ) space in a general pattern-forming system is often referred to as the Busse balloon or **stability balloon**. The Eckhaus, zigzag, cross-roll, skew-varicose and oscillatory instabilities, among others, can all define part of the boundary of the balloon. Sketches of typical Busse balloons for high and low Prandtl numbers are shown in Figure 8.13. In the high Prandtl number case a bimodal instability, where initial cross rolls lead to bimodal convection, forms part of the boundary.

8.7 The Benjamin–Feir instability

The Benjamin–Feir instability is an instability of travelling plane waves, analogous to the Eckhaus instability of rolls. In Chapter 7 we saw that slowly modulated

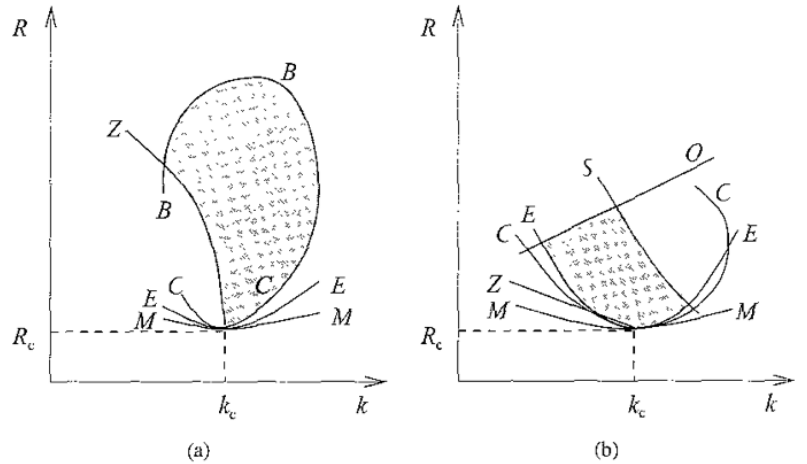


Fig 8 13 Sketches of a typical section through the Busse balloon at (a) high Prandtl number and (b) low Prandtl number The curve marked M is the marginal stability curve E, C, Z, B, S and O are the Eckhaus, cross-roll, zigzag, bimodal, skew-varicose and oscillatory instability boundaries respectively The region of stable rolls is shaded grey After Busse and Whitehead (1971) and Manneville (1990)

right-travelling waves in one spatial dimension can be written in the form

$$u(x,t) = A(Z,T)e^{i(t-x)} + c.c. + h.o.t., \tag{8 188}$$

where $Z = X - cT$ is a coordinate moving with the group velocity, c, and where the envelope, $A(Z, T)$, evolves according to the equation

$$\frac{\partial A}{\partial T} = \mu A + (1 + i\alpha)\frac{\partial^2 A}{\partial Z^2} - (1 + i\beta)|A|^2 A, \tag{8 189}$$

where μ is the bifurcation parameter and α and β are real constants Now, a plane right-travelling wave with wavenumber slightly different from critical takes the form

$$A = R_0 e^{i(\omega T - qX)} \equiv R_0 e^{i\{(\omega - qc)T - qZ\}}, \tag{8 190}$$

$$R_0^2 = (\mu - q^2), \tag{8 191}$$

$$\omega = cq - \alpha q^2 - \beta R_0^2 \tag{8 192}$$

Perturbing this solution so that

$$A = R_0(1 + r)e^{i\{(\omega - qc)T - qZ + \phi\}}, \tag{8 193}$$

where r and ϕ are real and $|r|, |\phi| \ll 1$, substituting into equation (8.189), separating real and imaginary parts and linearising in r and ϕ gives

$$\frac{\partial r}{\partial T} = \frac{\partial^2 r}{\partial Z^2} + 2q\frac{\partial \phi}{\partial Z} + 2\alpha q\frac{\partial r}{\partial Z} - \alpha\frac{\partial^2 \phi}{\partial Z^2} - 2R_0^2 r, \tag{8.194}$$

$$\frac{\partial \phi}{\partial T} = \alpha\frac{\partial^2 r}{\partial Z^2} + 2\alpha q\frac{\partial \phi}{\partial Z} - 2q\frac{\partial r}{\partial Z} + \frac{\partial^2 \phi}{\partial Z^2} - 2\beta R_0^2 r \tag{8.195}$$

Setting $r = \hat{r}e^{\sigma T + ikZ}$ and $\phi = \hat{\phi}e^{\sigma T + ikZ}$, where $\hat{r}$ and $\hat{\phi}$ are real constants, and where $|k| \ll |q|$, it is possible to deduce, after some calculation that in the long-wave limit, $|k| \to 0$, the two growth rate eigenvalues are

$$\sigma_1 = -2R_0^2 + O(k), \tag{8.196}$$

$$\sigma_2 = 2iq(\alpha - \beta)k - (1 + \alpha\beta)k^2 + \frac{2q^2k^2}{R_0^2}(1 + \beta^2) + O(k^3). \tag{8.197}$$

The first, as ever, represents the rapid relaxation of the amplitude to its equilibrium value, while the second corresponds to a phase mode. There will be instability for $Re(\sigma_2) > 0$, in other words for

$$-(1 + \alpha\beta) + \frac{2q^2}{R_0^2}(1 + \beta^2) > 0. \tag{8.198}$$

This will hold for all q, including the preferred mode $q = 0$ corresponding to waves at the critical wavelength, if

$$1 + \alpha\beta < 0. \tag{8.199}$$

The **Benjamin–Feir instability** sets in if this criterion is satisfied, and then uniform travelling waves of all wavelengths are unstable. It is a long-wave side-band instability, originally identified in wave trains on the free surface of a liquid: a wave of approximately constant wavelength and frequency breaks up, becoming highly irregular far from its source (Benjamin & Feir, 1967). In the original context the envelope equation takes the conservative form

$$\frac{\partial A}{\partial T} = i\alpha\frac{\partial^2 A}{\partial Z^2} - i\beta|A|^2 A, \tag{8.200}$$

which is the nonlinear Schrödinger equation, derived in Chapter 7, and the criterion for instability reduces to $\alpha\beta < 0$.

Figure 8.14 shows a numerical simulation of the Benjamin–Feir instability, where a uniform wavetrain breaks up into a number of unequal pulses.

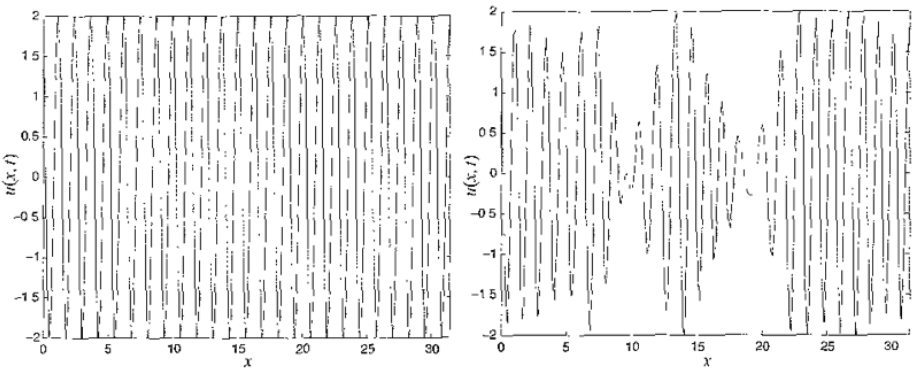


Fig. 8.14 The Benjamin–Feir instability in equation (8.189) with periodic boundary conditions, and $c = 0$, $\alpha = 1$, $\beta = -2$. The plots are of $u(x,t) = A(X,T)e^{i(t-x)} + c.c.$, where $\epsilon \equiv X/x = 0.2$. (a) The initial state at time $T = 0$ is a uniform travelling wave with $q = 0.0$ and $\mu = 1.0$ in the Benjamin–Feir unstable region. A small random perturbation has been added. The wavetrain breaks up into a series of uneven pulses as seen in (b) at time $T = 617$.

Returning to the $q = 0$ case of the full equations, we can expand the growth rate of the instability to fourth order in k to get

$$\sigma_2 = -(1 + \alpha\beta)k^2 - \frac{\alpha^2 k^4}{2R_0^2}(1 + \beta^2) + O(k^6) \tag{8.201}$$

So to quartic order the linear parts of the phase equation are

$$\frac{\partial \phi}{\partial T} = (1 + \alpha\beta)\frac{\partial^2 \phi}{\partial Z^2} - \frac{\alpha^2}{2R_0^2}(1 + \beta^2)\frac{\partial^4 \phi}{\partial Z^4} \tag{8.202}$$

Now the leading order nonlinear term in the phase gradient is $(\partial\phi/\partial Z)^2$. The analogous term, $(\partial\phi/\partial X)^2$, is forbidden in the phase equation for the Eckhaus instability as it does not satisfy x-reflection symmetry ($X \to -X$, $\phi \to -\phi$), but here there is no such symmetry as we are using a moving frame, and permitting only right-travelling waves. We can therefore include the term $(\partial\phi/\partial Z)^2$, giving the leading order phase equation

$$\frac{\partial \phi}{\partial T} = (1 + \alpha\beta)\frac{\partial^2 \phi}{\partial Z^2} - \frac{\alpha^2}{2R_0^2}(1 + \beta^2)\frac{\partial^4 \phi}{\partial Z^4} + g\left(\frac{\partial \phi}{\partial Z}\right)^2, \tag{8.203}$$

which is the **Kuramoto–Sivashinksy equation** (Kuramoto & Tsuzuki, 1976; Sivashinsky, 1977). The much-studied Kuramoto–Sivashinksy equation has

chaotic solutions, showing a transition from low-dimensional temporal chaos to **phase turbulence**, a chaotic regime characterised by spatiotemporal disorder.

Exercises

8.1 Assume you have a stripe solution with wavenumber $1+\epsilon q$, and that small phase perturbations, ϕ, to this solution satisfy the equation

$$\frac{\partial \phi}{\partial T} = D_{\parallel}(q)\frac{\partial^2 \phi}{\partial X^2} + D_{\perp}(q)\frac{\partial^2 \phi}{\partial Y^2} + f(q)\frac{\partial \phi}{\partial X}\frac{\partial^2 \phi}{\partial X^2} + g(q)\frac{\partial \phi}{\partial X}\frac{\partial^2 \phi}{\partial Y^2} + h(q)\frac{\partial \phi}{\partial Y}\frac{\partial^2 \phi}{\partial X \partial Y},$$

where $T = \epsilon^2 t$ is a slow timescale, $X = \epsilon x$ is a long scale in the across-axis direction and $Y = \epsilon^{1/2} y$ is a long scale in the along-axis direction as usual, and where $D_{\parallel}(q)$, $D_{\perp}(q)$, $f(q)$, $g(q)$ and $h(q)$ are real functions of q. Use Kuramoto's method to work out $f(q)$, $g(q)$ and $h(q)$ in terms of $D_{\parallel}(q)$ and $D_{\perp}(q)$.

8.2 Work out the cross-roll instability boundary for rolls governed by the Swift–Hohenberg equation. Does the cross-roll or the Eckhaus instability come in first?

8.3 What are the consequences of using free-slip boundary conditions, rather than periodic, for the Eckhaus instability in a finite domain?

8.4 If the phenomenological extension to no-slip boundaries of the analysis of the oscillatory instability of rolls gives a dispersion relation

$$\sigma^2 + \{\alpha + (a+c)k^2 + \beta k^4\}\sigma + bk^2 + \gamma k^4 = O(k^5),$$

where $\alpha > 0$, $\beta > 0$ and γ are real constants, show that there will be an oscillatory instability if $a + c$ is sufficiently negative. Work out a condition on γ for the instability to have wavenumber, $k_c \neq 0$, and frequency $\omega_c \neq 0$ at the critical value of $a + c$.

Now assume that the phase can be written in the form

$$\phi(Y, T) = \{A(\zeta, \tau)e^{i(\omega_c T + k_c Y)} + B(\zeta, \tau)e^{i(\omega_c T + k_c Y)} + c.c.\} + C(\zeta, \tau) + h.o.t.$$

in the nonlinear régime, where ζ and τ are long scales in the Y and T directions respectively and the functions $A(\zeta, \tau)$, $B(\zeta, \tau)$ and $C(\zeta, \tau)$ evolve according to the equations

$$\frac{\partial A}{\partial \tau} = \mu A - c_g\frac{\partial A}{\partial \zeta} + (a_1 - ia_2)\frac{\partial^2 A}{\partial \zeta^2} - (b_1 - ib_2)\{|A|^2 + 2|B|^2 + C^2\}A +$$

$$\frac{\partial B}{\partial \tau} = \mu B + c_g \frac{\partial B}{\partial \zeta} + (a_1 - ia_2)\frac{\partial^2 B}{\partial \zeta^2}$$
$$-(b_1 - ib_2)\{|B|^2 + 2|A|^2 + C^2\}B +$$
$$\frac{\partial C}{\partial \tau} = s\frac{\partial^2 C}{\partial \zeta^2},$$

where μ, c_g, $a_1 > 0$, a_2, b_1, b_2 and $s > 0$ are real. Considering spatially uniform solutions $A = Pe^{i\Omega\tau}$, $B = Qe^{i\Omega\tau}$, $C = 0$, where P and Q are complex constants, show that only propagating waves can be amplitude-stable.

Further details can be found in Fauve, Bolton and Brachet (1987)

9

More instabilities of patterns

Patterns such as hexagons and squares can also become unstable to phase and cross-pattern modes, while stripes in systems with additional symmetries, such as Galilean invariance, can undergo new types of instability, leading to drift, for example. This chapter looks at some of these new situations, starting with two examples of more complicated planforms – hexagons and quasipatterns – and some of their instabilities. After that we study drift instabilities where stationary or standing-wave patterns start to travel, and finally we look at the effect of Galilean invariance and conservation laws on the instabilities of stripes.

9.1 Instabilities of two-dimensional steady patterns

There are many possible extensions of the work on roll instabilities to more complicated situations. An obvious starting point is to consider what happens when the pattern that emerges at the primary pattern-forming instability is more complicated – a steady square pattern, for example, or oscillating hexagons, or maybe even a quasipattern. There is an extensive literature dealing with the phase instabilities of steady and oscillatory patterns of all sorts. We shall concentrate on two examples – steady hexagons and steady twelvefold quasipatterns – that illustrate how to extend the methods used in the previous chapter to these harder problems and lead to some interesting new results. At the end of this chapter you will find exercises on the instabilities of steady and oscillating squares as further examples.

9.1.1 Instabilities of hexagons

In this section we will adapt the methods used for rolls to investigate the instabilities of hexagons. Many researchers have worked on this problem; further details can be found in Hoyle (1994a, 1995, 1998b), where the exposition is similar to the one I shall present here, and in Echebarria and Pérez-García (1998), Kuske and

Milewski (1999) and many other papers, several of which are referenced in these five.

Consider a hexagonal pattern written

$$u(x, y, t) = A(X, Y, T)e^{ix} + B(X, Y, T)e^{i(-x+\sqrt{3}y)/2} + C(X, Y, T)e^{-i(x+\sqrt{3}y)/2} + c.c. + h.o.t. \tag{9.1}$$

For simplicity we will consider the case where there are no spatial derivatives in the quadratic terms in the amplitude equations; as discussed in Section 7.7, this is often reasonable. Close to the onset of the instability, the evolution of the pattern can then be described to leading order by the three coupled nonlinear equations

$$\frac{\partial A}{\partial T} = \mu A + a\bar{B}\bar{C} - b|A|^2A - c(|B|^2 + |C|^2)A + \frac{\partial^2 A}{\partial X_1^2}, \tag{9.2}$$

$$\frac{\partial B}{\partial T} = \mu B + a\bar{A}\bar{C} - b|B|^2B - c(|C|^2 + |A|^2)B + \frac{\partial^2 B}{\partial X_2^2}, \tag{9.3}$$

$$\frac{\partial C}{\partial T} = \mu C + a\bar{A}\bar{B} - b|C|^2C - c(|A|^2 + |B|^2)C + \frac{\partial^2 C}{\partial X_3^2}, \tag{9.4}$$

where $X_1 = X$, $X_2 = (-X + \sqrt{3}Y)/2$ and $X_3 = -(X + \sqrt{3}Y)/2$, and μ, a, b and c are real constants. These are equations (7.116)–(7.118), derived in Section 7.7, with $g = h = 0$. If you are interested in what happens if g and h are nonzero, so that spatial-derivative quadratic terms are included, you can find the relevant calculations in Echebarria and Pérez-García (1998) and Kuske and Milewski (1999).

A perfect hexagonal pattern at slightly off-critical wavenumber is given by $A = R_0e^{iqX_1}$, $B = R_0e^{iqX_2}$, $C = R_0e^{iqX_3}$, where $R_0 > 0$ and q are real constants. Substituting into equations (9.2)–(9.4) gives

$$R_0^2(b + 2c) - aR_0 - (\mu - q^2) = 0 \tag{9.5}$$

In this case the hexagon phase, $\Phi = \mathrm{Arg}(ABC)$, is zero, and so these are up-hexagons. We could also choose $\Phi = \pi$ to give down-hexagons, for example by setting $A = -R_0e^{iqX_1}$, $B = R_0e^{iqX_2}$ and $C = R_0e^{iqX_3}$, where

$$R_0^2(b + 2c) + aR_0 - (\mu - q^2) = 0 \tag{9.6}$$

Allowing small amplitude and phase perturbations to up-hexagons of the form

$$A = R_0e^{i(qX_1+\phi_1)}(1 + r_1), \tag{9.7}$$

$$B = R_0e^{i(qX_2+\phi_2)}(1 + r_2), \tag{9.8}$$

$$C = R_0e^{i(qX_3+\phi_3)}(1 + r_3), \tag{9.9}$$

where $|r_i(X_1, X_2, X_3, T)| \ll 1$ and $|\phi_i(X_1, X_2, X_3, T)| \ll 1$ for $i = 1, 2, 3$, and linearising in the r_i and ϕ_i results in the equations

$$\frac{\partial r_1}{\partial T} = aR_0(-r_1 + r_2 + r_3) - 2bR_0^2 r_1 - 2cR_0^2(r_2 + r_3) - 2q\frac{\partial \phi_1}{\partial X_1} + \frac{\partial^2 r_1}{\partial X_1^2}, \tag{9.10}$$

$$\frac{\partial \phi_1}{\partial T} = -aR_0(\phi_1 + \phi_2 + \phi_3) + 2q\frac{\partial r_1}{\partial X_1} + \frac{\partial^2 \phi_1}{\partial X_1^2}, \tag{9.11}$$

with four other equations obtained by permuting the subscripts $\{1, 2, 3\}$ cyclically. In the absence of spatial modulations ($\partial/\partial X_1 = \partial/\partial X_2 = \partial/\partial X_3 = 0$) the perturbations evolve according to

$$\frac{\mathrm{d}}{\mathrm{d}T}(r_1 + r_2 + r_3) = -v(r_1 + r_2 + r_3), \tag{9.12}$$

$$\frac{\mathrm{d}}{\mathrm{d}T}(r_i - r_j) = -2u(r_i - r_j), \quad i \neq j \tag{9.13}$$

$$\frac{\mathrm{d}}{\mathrm{d}T}(\phi_1 + \phi_2 + \phi_3) = -3aR_0(\phi_1 + \phi_2 + \phi_3), \tag{9.14}$$

$$\frac{\mathrm{d}}{\mathrm{d}T}(\phi_i - \phi_j) = 0, \quad i \neq j \tag{9.15}$$

where

$$u = R_0^2(b - c) + aR_0, \tag{9.16}$$

$$v = 2R_0^2(b + 2c) - aR_0. \tag{9.17}$$

These equations are in agreement with those found in Section 5.4 for the stability of hexagons on a hexagonal lattice. The amplitude and phase perturbations correspond to the real and imaginary parts respectively of the perturbations δz_i used there. Without loss of generality we can take $a > 0$, since if $a < 0$ we can apply the transformation $a \to -a$, $\Phi \to \Phi + \pi$, and recover the equations (9.10)–(9.15). The hexagons are therefore stable to perturbations in the hexagon phase $\Phi = \phi_1 + \phi_2 + \phi_3$. The remaining two phase modes, which are associated with translations in two horizontal directions, are neutral in the absence of spatial modulation; these are the true phase modes. If we also choose μ, a, b and c such that $u > 0$ and $v > 0$, then the pattern is stable at leading order to disturbances in the amplitudes. The inequalities $u > 0$ and $v > 0$ require that $b + c > 0$, and we also have $b + 2c > 0$ from $v > 0$ and $a > 0$. Consequently, if $b - c < 0$, the relevant parameter regime is $-c < b < c$, with $c > 0$, whereas if $b - c > 0$, it is $-b/2 < c < b$, with $b > 0$.

When spatial modulation is allowed, the two phase modes are no longer neutral, but evolve on long space and time scales. Using the phase approximation, where the stable amplitude modes are slaved adiabatically to the unstable phase modes in the long-wave limit $\partial/\partial X_1, \partial/\partial X_2, \partial/\partial X_3 \sim \delta$, $|\delta| \ll 1$, the following phase equations can be derived at leading order:

$$\frac{\partial}{\partial T}(\phi_2+\phi_3) = \left\{-\frac{q^2}{2u}\nabla^2 + \frac{1}{4}\left(3\frac{\partial^2}{\partial X^2}+\frac{\partial^2}{\partial Y^2}\right) - \frac{2q^2}{v}\frac{\partial^2}{\partial X^2}\right\}(\phi_2+\phi_3) - \frac{1}{2\sqrt{3}}\left(1-\frac{4q^2}{v}\right)\frac{\partial^2}{\partial X\partial Y}(\phi_2-\phi_3), \tag{9.18}$$

$$\frac{\partial}{\partial T}(\phi_2-\phi_3) = \left\{-\frac{q^2}{2u}\nabla^2 + \frac{1}{4}\left(\frac{\partial^2}{\partial X^2}+3\frac{\partial^2}{\partial Y^2}\right) - \frac{2q^2}{v}\frac{\partial^2}{\partial Y^2}\right\}(\phi_2-\phi_3) - \frac{\sqrt{3}}{2}\left(1-\frac{4q^2}{v}\right)\frac{\partial^2}{\partial X\partial Y}(\phi_2+\phi_3) \tag{9.19}$$

Setting $\phi_2 = \widehat{\phi}_2 e^{\sigma T+ikX+ilY}$ and $\phi_3 = \widehat{\phi}_3 e^{\sigma T+ikX+ilY}$, we find the growth rate eigenvalues

$$\sigma_1 \equiv \frac{1}{4}\left(\frac{2q^2}{u}-1\right)(k^2+l^2)+O(k^3), \tag{9.20}$$

$$\sigma_2 \equiv \frac{1}{4}\left(\frac{2q^2}{u}+\frac{8q^2}{v}-3\right)(k^2+l^2)+O(k^3) \tag{9.21}$$

The phase mode corresponding to σ_1 satisfies

$$k(\phi_2+\phi_3) = \frac{l}{\sqrt{3}}(\phi_2-\phi_3). \tag{9.22}$$

Since the hexagon phase perturbation is zero to leading order ($\phi_1+\phi_2+\phi_3=0$), because it is slaved to the two true phase modes, equation (9.22) is equivalent to

$$\frac{\partial\phi_x}{\partial X} = -\frac{\partial\phi_y}{\partial Y}, \tag{9.23}$$

where $\phi_x = \phi_1$ is the x phase, and $\phi_y = (\phi_2-\phi_3)/\sqrt{3}$ is the y phase. (Note that the translation $x \to x+x_0$, $y \to y+y_0$, gives $\phi_1 \to \phi_1 - x_0$, $\phi_2-\phi_3 \to \phi_2-\phi_3-\sqrt{3}y_0$, so we can identify the x and y phases as claimed.) This represents stretching along one Cartesian axis in wavevector space and contraction along the other, and so we call this the rectangular Eckhaus instability of hexagons. If we write $\boldsymbol{\phi} = (\phi_x, \phi_y)$, equation (9.23) takes the form $\nabla \cdot \boldsymbol{\phi} = 0$, so the phase perturbation is divergence-free. Some tilting of the Cartesian grid may also occur, since $\partial\phi_x/\partial Y$ and $\partial\phi_y/\partial X$ are not restricted. The evolution of the instability is

shown in Figure 9.1: an initial pattern with a short wavelength (large q) in the rectangular Eckhaus unstable band evolves towards a longer wavelength

Close to the phase stability boundary, we can set $q^2 = q_0^2 + d\delta^2$, with $|\delta| \ll 1$ and d measuring the distance from threshold, and perform a weakly nonlinear analysis. The critical wavenumber, q_0, is defined by $2q_0^2/u = 1$. After a lengthy calculation, the leading order phase equation for the rectangular instability turns out to be

$$\begin{aligned}\frac{\partial}{\partial T}\nabla^2\phi_1 = &-\frac{d}{4q_0^2}\nabla^4\phi_1 - \frac{1}{64q_0^2}\left(\nabla^6\phi_1 - 8\frac{\partial^6\phi_1}{\partial X_1^2\partial X_2^2\partial X_3^2}\right)\\ &+\frac{(6q_0^2 - aR_0)}{6q_0^3}\frac{\partial^2}{\partial X_2\partial X_3}\left(\frac{\partial}{\partial X_3} - \frac{\partial}{\partial X_2}\right)\left\{\left(\frac{\partial\phi_2}{\partial X_2}\right)^2 - \left(\frac{\partial\phi_3}{\partial X_3}\right)^2\right\}\\ &-\frac{(6q_0^2 - aR_0)}{18q_0^3}\left(\frac{\partial}{\partial X_3} - \frac{\partial}{\partial X_2}\right)^2\left\{\frac{\partial}{\partial X_1}\left(\frac{\partial\phi_1}{\partial X_1}\right)^2\right.\\ &\left.+\frac{\partial}{\partial X_2}\left(\frac{\partial\phi_2}{\partial X_2}\right)^2 + \frac{\partial}{\partial X_3}\left(\frac{\partial\phi_3}{\partial X_3}\right)^2\right\}.\end{aligned} \tag{9.24}$$

This equation describes the evolution of the Laplacian of the phase; the evolution of the phase itself will depend upon the conditions at the lateral boundaries, so the problem is nonlocal. The bifurcation can be shown to be subcritical (Hoyle, 1994a, 1995), leading to the breakdown of the pattern through the creation of **penta-hepta** defects, where two hexagons are replaced by a pentagon and a heptagon, as can be seen in Figure 9.1.

The phase mode corresponding to σ_2, defined in equation (9.21), satisfies

$$l(\phi_2 + \phi_3) = -\frac{k}{\sqrt{3}}(\phi_2 - \phi_3) \tag{9.25}$$

at leading order, which can be rewritten

$$\frac{\partial\phi_x}{\partial Y} = \frac{\partial\phi_y}{\partial X}, \tag{9.26}$$

or $\nabla \times \phi = 0$, so the phase perturbation is irrotational. If the original hexagons were drawn on a sheet of rubber, this phase mode has the effect of rotating Cartesian axes in the rubber towards or away from each other. We call this mode rhombic, because it would turn a square grid aligned along the Cartesian axes into a rhombic lattice. The pattern may also expand or contract along the coordinate axes because $\partial\phi_x/\partial X$ and $\partial\phi_y/\partial Y$ are not restricted.

Close to the rhombic Eckhaus instability boundary we can set $q^2 = q_0^2 + d\delta^2$, where $|\delta| \ll 1$ and d measures the distance from threshold, and where the critical wavenumber, q_0, is now defined by $2q_0^2/u + 8q_0^2/v = 3$, and once more carry out a weakly nonlinear analysis. After an extremely long and somewhat painful

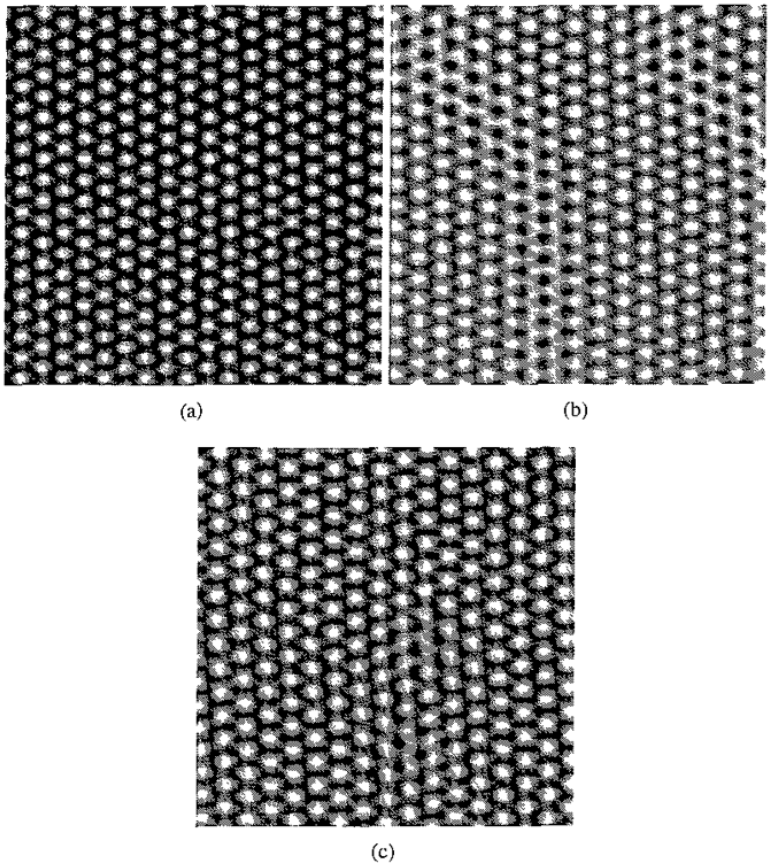


Fig 9 1 Greyscale plots of $u = A(X, Y, T)e^{ix} + B(X, Y, T)e^{i(-x+\sqrt{3}y)/2} + C(X, Y, T)e^{-i(x+\sqrt{3}y)/2} + c\,c$ for a hexagonal pattern undergoing the rectangular Eckhaus instability during numerical integration of the amplitude equations (9.2–9.4). The parameter values are $q = 9\,0$, $\mu = 461\,0$, $\epsilon = 0\,125$, $a = 2\,0$, $b = 2\,0$ and $c = 1.0$ The plots are shown at times (*a*) $T = 0.0$, (*b*) $T = 13\,0$ and (*c*) $T = 30\,0$ Only one-hundredth of the original integration domain is shown in each case, and the region illustrated is chosen in each case to show the most interesting behaviour, so it is different for each of the three plots

calculation, the phase equation for the rhombic Eckhaus instability turns out to be

$$\frac{\partial}{\partial T}\nabla^2\phi_1 = -\frac{3d}{4q_0^2}\nabla^4\phi_1 - \frac{9}{64q_0^2}\nabla^6\phi_1 - \frac{2q_0^2}{u^2}\frac{\partial^6\phi_1}{\partial X_1^2\partial X_2^2\partial X_3^2} + A\left\{\frac{\partial^2}{\partial X_1^2}\left(\frac{\partial\phi_1}{\partial X_1}\Psi\right) + \frac{\partial^2}{\partial X_1\partial X_2}\left(\frac{\partial\phi_2}{\partial X_2}\Psi\right) + \frac{\partial^2}{\partial X_1\partial X_3}\left(\frac{\partial\phi_3}{\partial X_3}\Psi\right)\right\}$$

$$+B\frac{\partial}{\partial X_1}\nabla^2\Theta^2+C\frac{\partial}{\partial X_1}\nabla^2\left\{\left(\frac{\partial\phi_1}{\partial X_1}\right)^2+\left(\frac{\partial\phi_2}{\partial X_2}\right)^2+\left(\frac{\partial\phi_3}{\partial X_3}\right)^2\right\}$$
$$+D\frac{\partial}{\partial X_1}\left\{\frac{\partial^2}{\partial X_1^2}\left(\frac{\partial\phi_1}{\partial X_1}\Theta\right)+\frac{\partial^2}{\partial X_2^2}\left(\frac{\partial\phi_2}{\partial X_2}\Theta\right)+\frac{\partial^2}{\partial X_3^2}\left(\frac{\partial\phi_3}{\partial X_3}\Theta\right)\right\}$$
$$+E\frac{\partial}{\partial X_1}\left\{\frac{\partial}{\partial X_1}\left(\frac{\partial\phi_1}{\partial X_1}\frac{\partial\Theta}{\partial X_1}\right)+\frac{\partial}{\partial X_2}\left(\frac{\partial\phi_2}{\partial X_2}\frac{\partial\Theta}{\partial X_2}\right)+\frac{\partial}{\partial X_3}\left(\frac{\partial\phi_3}{\partial X_3}\frac{\partial\Theta}{\partial X_3}\right)\right\}$$
$$+F\frac{\partial}{\partial X_1}\left\{\frac{\partial^2}{\partial X_1^2}\left(\frac{\partial\phi_2}{\partial X_2}\frac{\partial\phi_3}{\partial X_3}\right)+\frac{\partial^2}{\partial X_2^2}\left(\frac{\partial\phi_3}{\partial X_3}\frac{\partial\phi_1}{\partial X_1}\right)+\frac{\partial^2}{\partial X_3^2}\left(\frac{\partial\phi_1}{\partial X_1}\frac{\partial\phi_2}{\partial X_2}\right)\right\}$$
$$+G\frac{\partial}{\partial X_1}\left\{\frac{\partial^2}{\partial X_1^2}\left(\frac{\partial\phi_1}{\partial X_1}\right)^2+\frac{\partial^2}{\partial X_2^2}\left(\frac{\partial\phi_2}{\partial X_2}\right)^2+\frac{\partial^2}{\partial X_3^2}\left(\frac{\partial\phi_3}{\partial X_3}\right)^2\right\}, \tag{9.27}$$

where

$$\Psi=\frac{\partial^2\phi_1}{\partial X_1^2}+\frac{\partial^2\phi_2}{\partial X_2^2}+\frac{\partial^2\phi_3}{\partial X_3^2}, \tag{9.28}$$

$$\Theta=\frac{\partial\phi_1}{\partial X_1}+\frac{\partial\phi_2}{\partial X_2}+\frac{\partial\phi_3}{\partial X_3}, \tag{9.29}$$

and where $A-G$ are functions of q_0, u, a and R_0, the exact forms being given in Hoyle (1994a, 1995)

Again the equation is for the Laplacian of the phase, so the evolution of the phase itself depends on the lateral boundary conditions Once more it can be shown that the bifurcation is subcritical, leading to the formation of defects, as seen in Figure 9 2. The figure also shows that the wavelength remains almost constant throughout the course of the rhombic instability. This is not surprising, since the rhombic phase relation involves neither $\partial\phi_x/\partial X$ nor $\partial\phi_y/\partial Y$, and so no stretching or compression is required However, the pattern is tilted, and shows bands of up-hexagons (where the hexagon phase, Ψ, is zero), down-hexagons ($\Psi=\pi$) and triangles ($\Psi=\pi/2$) It appears that the rhombic instability stabilises the pattern by introducing local variations in the hexagon phase, rather than by changing its wavelength

In the most commonly observed situation (see Busse, 1967) where hexagons are the stable solution at small μ and rolls are stable at large μ, the coefficients of the amplitude equations satisfy $b-c<0$ and $b>0$. In this case the region of stability to amplitude and phase modes is closed, as shown in Figure 9.3, in agreement with numerical calculations of the stability balloon (see for example, Iveitereid & Palm, 1976; Bestehorn, 1994) This contrasts with the roll case, where the region of stability to phase modes is open in the weakly nonlinear approximation where the envelope equation framework is valid In Figure 9.3 the rhombic instability sets

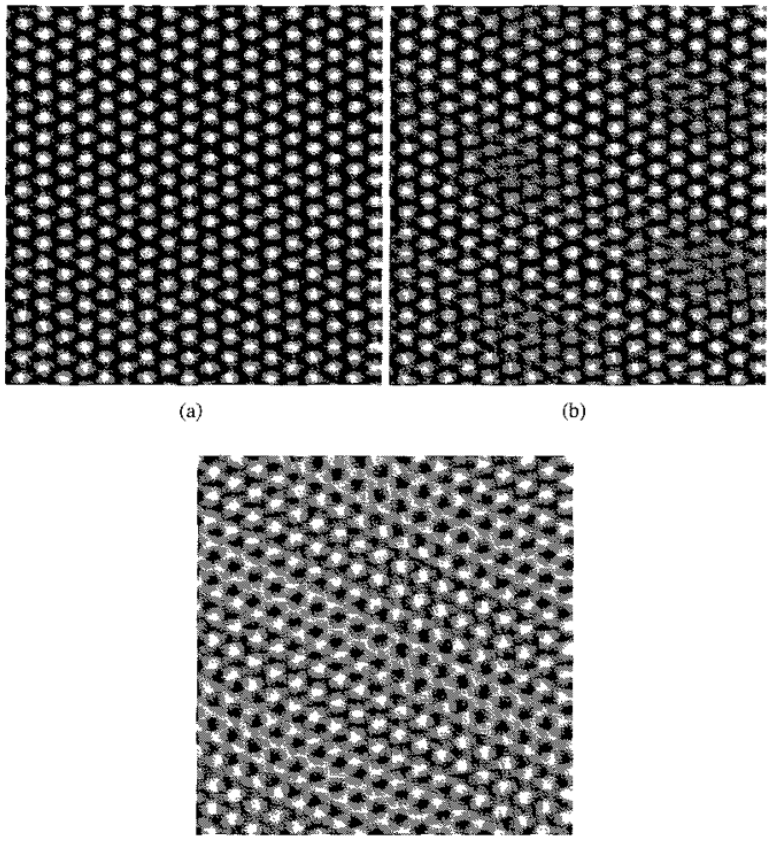


Fig 9 2 Greyscale plots of $u = A(X, Y, T)e^{ix} + B(X, Y, T)e^{i(-x+\sqrt{3}y)/2} + C(X, Y, T)e^{-i(x+\sqrt{3}y)/2} + c\,c$ for a hexagonal pattern undergoing the rhombic Eckhaus instability during numerical integration of the amplitude equations (9 2–9 4). The parameter values are $q = 12\,0$, $\mu = 224.0$, $\epsilon = 0.125$, $a = 2\,0$, $b = 3\,0$ and $c = -1.0$ The plots are shown at times (*a*) $T = 0.00$, (*b*) $T = 0.14$ and (*c*) $T = 0\,20$. The bottom-left one-hundredth of the original integration domain is shown in each case From Hoyle (1994a)

in first for $\mu < a^2(3b + c)/8c^2$, while for $\mu > a^2(3b + c)/8c^2$ it is the rectangular mode that defines the edge of the stable region. In fact, the two instability boundaries always cross except in the parameter regime $b - c > 0$ and $c < 0$, when the rhombic instability is the more dangerous for all values of μ

Numerical simulations (Sushchik & Tsimring, 1994; Bestehorn, 1993) show that the Eckhaus instabilities can trigger the hexagon-roll transition when μ is close to the value, μ_s, at which a front between rolls and hexagons would be stationary. When $\mu = \mu_s$, both rolls and hexagons at critical wavelength are stable,

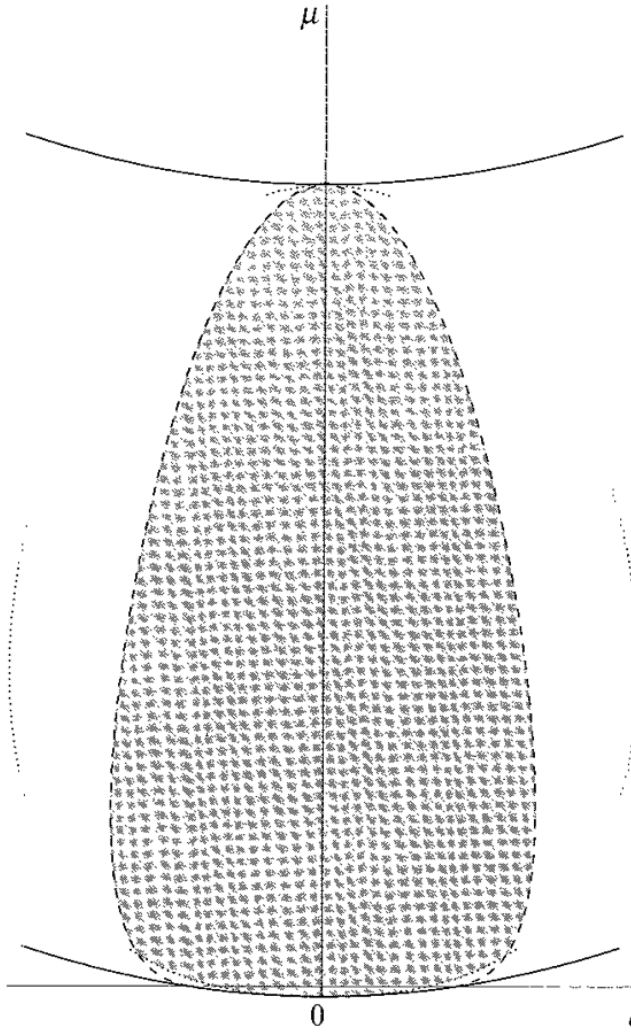


Fig 9 3 Hexagon phase stability boundaries for $b-c<0$, $b>0$. The dotted and dashed curves are the rhombic and rectangular Eckhaus stability boundaries respectively The two curves cross at $q=\pm(a/2c)\sqrt{(b+c)/2}$, $\mu=a^2(3b+c)/8c^2$ $(R_0=a/2c)$. The area of stable hexagons is shaded grey The curves $\mu=q^2-a^2/4(b+2c)$ (lower curve) and $\mu=q^2+a^2(2b+c)/(b-c)^2$ (upper curve) are shown as solid lines These correspond to $R_0=a/2(b+2c)$ and $R_0=-a/(b-c)$ respectively. For $\mu<q^2-a^2/4(b+2c)$ no hexagons exist, while for $\mu>q^2+a^2(2b+c)/(b-c)^2$ hexagons are unstable to rolls (see Section 5 4 and Figure 5 8 for further details)

and their free energies are equal, as we will see in Chapter 10. In fact, Sushchik and Tsimring (1994) observe that the hexagon to roll transition is initiated by pieces of roll in the cores of penta-hepta defects. There is also evidence that the Eckhaus, cross-roll and zigzag instabilities of rolls can trigger the transition to hexagons (Lauzeral, Metens & Walgraef, 1993; Bestehorn, 1993; Sushchik & Tsimring, 1994).

There is no zigzag or two-dimensional Eckhaus instability of hexagons. By analogy with rolls, a zigzag phase disturbance to the hexagonal pattern would take the form $\phi_i \equiv \phi_i(Y_i, T)$, $i = 1, 2, 3$, where $Y_1 = Y$, $Y_2 = -\sqrt{3}X/2 - Y/2$, $Y_3 = \sqrt{3}X/2 - Y/2$. To leading order, the phase-locking equation

$$\phi_1(Y_1, T) + \phi_2(Y_2, T) + \phi_3(Y_3, T) = 0 \tag{9.30}$$

holds. Differentiating twice with respect to $Y = Y_1$ gives

$$\frac{\partial^2 \phi_1}{\partial Y_1^2} + \frac{1}{4}\left(\frac{\partial^2 \phi_2}{\partial Y_2^2} + \frac{\partial^2 \phi_3}{\partial Y_3^2}\right) = 0. \tag{9.31}$$

Two further equations obtained by cyclically permuting the subscripts $\{1, 2, 3\}$ also hold. Combining all three gives

$$\frac{\partial^2 \phi_1}{\partial Y_1^2} = \frac{\partial^2 \phi_2}{\partial Y_2^2} = \frac{\partial^2 \phi_3}{\partial Y_3^2} = 0 \tag{9.32}$$

So each phase disturbance ϕ_i can be at most a linear function of the corresponding Y_i. It can be seen from equations (9.18) and (9.19) that such perturbations do not grow, and so there is no zigzag instability of hexagons. This has also been demonstrated numerically (Sushchik & Tsimring 1994). Similarly, a two-dimensional Eckhaus phase disturbance would take the form $\phi_i \equiv \phi_i(X_i, T)$, $i = 1, 2, 3$, where

$$\frac{\partial^2 \phi_1}{\partial X_1^2} = \frac{\partial^2 \phi_2}{\partial X_2^2} = \frac{\partial^2 \phi_3}{\partial X_3^2} = 0. \tag{9.33}$$

Again these perturbations do not grow and so there is no two-dimensional Eckhaus instability of hexagons.

We can also generalise the cross-roll instability to the hexagon case. A cross-hexagon perturbation would take the form of a hexagonal pattern growing at an angle, θ, to the original hexagons. Symmetry considerations suggest that the six component roll amplitudes satisfy the equations

$$\begin{aligned}\frac{\partial A}{\partial T} = {} & \mu A + a\bar{B}\bar{C} - b|A|^2 A - c(|B|^2 + |C|^2)A \\ & - d_1|\widehat{A}|^2 A - d_2|\widehat{B}|^2 A - d_3|\widehat{C}|^2 A + \frac{\partial^2 A}{\partial X_1^2},\end{aligned} \tag{9.34}$$

$$\frac{\partial B}{\partial T} = \mu B + a\bar{A}\bar{C} - b|B|^2 B - c(|A|^2 + |C|^2)B$$
$$-d_1|\widehat{B}|^2 B - d_2|\widehat{C}|^2 B - d_3|\widehat{A}|^2 B + \frac{\partial^2 B}{\partial X_2^2}, \tag{9.35}$$

$$\frac{\partial C}{\partial T} = \mu C + a\bar{A}\bar{B} - b|C|^2 C - c(|A|^2 + |B|^2)C$$
$$-d_1|\widehat{C}|^2 C - d_2|\widehat{A}|^2 C - d_3|\widehat{B}|^2 C + \frac{\partial^2 C}{\partial X_3^2}, \tag{9.36}$$

$$\frac{\partial \widehat{A}}{\partial T} = \mu \widehat{A} + a\bar{\widehat{B}}\bar{\widehat{C}} - b|\widehat{A}|^2 \widehat{A} - c(|\widehat{B}|^2 + |\widehat{C}|^2)\widehat{A}$$
$$-d_1|A|^2 \widehat{A} - d_2|C|^2 \widehat{A} - d_3|B|^2 \widehat{A} + \frac{\partial^2 \widehat{A}}{\partial \widehat{X}_1^2}, \tag{9.37}$$

$$\frac{\partial \widehat{B}}{\partial T} = \mu \widehat{B} + a\bar{\widehat{A}}\bar{\widehat{C}} - b|\widehat{B}|^2 \widehat{B} - c(|\widehat{A}|^2 + |\widehat{C}|^2)\widehat{B}$$
$$-d_1|B|^2 \widehat{B} - d_2|A|^2 \widehat{B} - d_3|C|^2 \widehat{B} + \frac{\partial^2 \widehat{B}}{\partial \widehat{X}_2^2}, \tag{9.38}$$

$$\frac{\partial \widehat{C}}{\partial T} = \mu \widehat{C} + a\bar{\widehat{A}}\bar{\widehat{B}} - b|\widehat{C}|^2 \widehat{C} - c(|\widehat{A}|^2 + |\widehat{B}|^2)\widehat{C}$$
$$-d_1|C|^2 \widehat{C} - d_2|B|^2 \widehat{C} - d_3|A|^2 \widehat{C} + \frac{\partial^2 \widehat{C}}{\partial \widehat{X}_3^2}, \tag{9.39}$$

where $\widehat{X}_1$, $\widehat{X}_2$ and $\widehat{X}_3$ are related in the same way as X_1, X_2 and X_3, and where there is an angle, θ, between the X_1 and $\widehat{X}_1$ axes. If all the constituent rolls were at critical wavelength then the general solution

$$u = Ae^{ix} + Be^{i(-x+\sqrt{3}y)/2} + Ce^{-i(x+\sqrt{3}y)/2}$$
$$+\widehat{A}e^{i\widehat{x}} + \widehat{B}e^{i(-\widehat{x}+\sqrt{3}\widehat{y})/2} + \widehat{C}e^{-i(\widehat{x}+\sqrt{3}\widehat{y})/2} + c.c., \tag{9.40}$$

where the $\widehat{x}$ and $\widehat{y}$ variables are related to $\widehat{X}$ and $\widehat{Y}$ in the obvious manner, would be doubly periodic on a superlattice for a countably infinite set of values of the angle θ (Dionne, Silber & Skeldon, 1997), (see Chapter 6) For other values of θ, the system describes quasipatterns, and the use of the amplitude equations cannot be justified rigorously, even for patterns at critical wavelength, much less when modulation is present

Assuming that the amplitude equations hold, we start with a hexagonal pattern $A = R_0 e^{iqX_1}$, $B = R_0 e^{iqX_2}$, $C = R_0 e^{iqX_3}$, and make a perturbation, $\widehat{A} = \widetilde{A}(T)e^{i\widehat{X}_1}$, $\widehat{B} = \widetilde{B}(T)e^{i\widehat{X}_2}$, $\widehat{C}(T) = \widetilde{C}e^{it\widehat{X}_3}$ where $|\widetilde{A}|$, $|\widetilde{B}|$, $|\widetilde{C}| \ll R_0$

Linearising, we have

$$\frac{d\widetilde{A}}{dT} = \{\mu - r^2 - (d_1 + d_2 + d_3)R_0^2\}\widetilde{A}, \tag{9.41}$$

$$\frac{d\widetilde{B}}{dT} = \{\mu - s^2 - (d_1 + d_2 + d_3)R_0^2\}\widetilde{B}, \tag{9.42}$$

$$\frac{d\widetilde{C}}{dT} = \{\mu - t^2 - (d_1 + d_2 + d_3)R_0^2\}\widetilde{C}. \tag{9.43}$$

The maximum growth rate of the perturbation occurs at the critical wavenumber $r = s = t = 0$, and is given by $\sigma = \mu - (d_1 + d_2 + d_3)R_0^2$. The disturbances are linearly uncoupled, and so it is not until the nonlinear stage that a hexagonal pattern would emerge amongst them.

The cross-hexagon boundary is given by

$$\mu = (d_1 + d_2 + d_3)R_0^2, \tag{9.44}$$

or equivalently,

$$q^2 = aR_0 + (d_1 + d_2 + d_3 - b - 2c)R_0^2 \tag{9.45}$$

Sushchik and Tsimring (1994) considered the case $d_1 = d_2 = d_3 = c$ where the coupling between different sets of rolls is independent of the angle between them, and found that in this case the cross-hexagon mode was never the primary means of breakdown of the pattern. For other choices of the parameters, the cross-hexagon mode can be the dominant one, marking the boundary of the region of stable hexagons for a range of μ.

White (1988) performed experiments on hexagonal cells in convecting Lyle's Golden Syrup. He found three types of instability that alter the pattern wavelength: mosaic, where small hexagons grow at the corners of the original cells, cell fusion and cell splitting. The initial number of cells was relatively small, however, so these are unlikely to be long-wave phase modes: cell fusion, for example, is probably a period-multiplying bifurcation of the type discussed in Chapter 6 and the mosaic instability looks a bit like a cross-hexagon mode. White also found similar instabilities of squares.

9.1.2 Phase and phason instabilities of quasipatterns

Similar methods can be used to study the phase instabilities of quasipatterns, though in the eyes of the purist using a set of two-dimensional modulation equations to describe the instabilities of patterns that can't be described properly by amplitude equations in the first place must surely be adding insult to injury. So much for purity! Let's press on undeterred, and see what happens.

Echebarria and Riecke (2001) have worked out the phase equations for octagonal (eightfold), decagonal (tenfold) and dodecagonal (twelvefold) quasipatterns In the twelvefold case, the pattern is written

$$u(x, y, t) = \sum_{n=1}^{6} A_n(X, Y, T)e^{ik_n x} + c.c. \tag{9.46}$$

to leading order, where the k_i are given in Figure 6 10d and satisfy $|k_i| = 1$, and where $X = \epsilon x$, $Y = \epsilon y$ and $T = \epsilon^2 t$ for $0 < \epsilon \ll 1$ To cubic order the A_n evolve according to the equations

$$\frac{\partial A_i}{\partial T} = \mu A_i + (k_i \cdot \nabla)^2 A_i + \alpha \bar{A}_{i+2}\bar{A}_{i+4} - A_i\{\beta|A_i|^2 + \delta(|A_{i+2}|^2 + |A_{i+4}|^2) + \gamma(|A_{i-1}|^2 + |A_{i-3}|^2) + \nu|A_{i+1}|^2\}, \quad i = 1, 3, 5, \tag{9.47}$$

$$\frac{\partial A_i}{\partial T} = \mu A_i + (k_i \cdot \nabla)^2 A_i + \alpha \bar{A}_{i+2}\bar{A}_{i+4} - A_i\{\beta|A_i|^2 + \delta(|A_{i+2}|^2 + |A_{i+4}|^2) + \gamma(|A_{i+1}|^2 + |A_{i+3}|^2) + \nu|A_{i-1}|^2\}, \quad i = 2, 4, 6, \tag{9.48}$$

where the indices cycle with period 6, where $\nabla = (\partial/\partial X, \partial/\partial Y)$, and where μ is the bifurcation parameter and α, β, δ, γ and ν are real constants. These equations are just the same as (6 53) and (6.54) with the addition of the modulation terms $(k_i \cdot \nabla)^2 A_i$.

A perfect dodecagonal quasipattern at slightly off-critical wavenumber is given by $A_j = R_0 e^{iqk_j \cdot X}$, where R_0 and q are real constants that satisfy

$$\mu - q^2 + \alpha R_0 - (\beta + 2\delta + 2\gamma + \nu)R_0^2 = 0. \tag{9.49}$$

This solution is then perturbed so that

$$A_j(X, Y, T) = R_0(1 + r_j(X, Y, T))e^{i(qk_j \cdot X + \phi_j(X, Y, T))}, \tag{9.50}$$

for $j = 1, \ldots, 6$. The amplitude perturbations, r_j, and the two global phases

$$\Phi_1 = \phi_1 + \phi_3 + \phi_5, \tag{9.51}$$

$$\Phi_2 = \phi_2 + \phi_4 + \phi_6, \tag{9.52}$$

are determined adiabatically by the two phase modes

$$\phi_x = \frac{1}{3}\left(\phi_1 - \frac{1}{2}\phi_3 - \frac{\sqrt{3}}{2}\phi_4 - \frac{1}{2}\phi_5 + \frac{\sqrt{3}}{2}\phi_6\right), \tag{9.53}$$

$$\phi_y = \frac{1}{3}\left(\phi_2 + \frac{\sqrt{3}}{2}\phi_3 - \frac{1}{2}\phi_4 - \frac{\sqrt{3}}{2}\phi_5 - \frac{1}{2}\phi_6\right), \tag{9.54}$$

corresponding to translations in the x and y directions respectively, and the two phason modes

$$\psi_x = \frac{2}{3}\left(\phi_1 - \frac{1}{2}\phi_3 + \frac{\sqrt{3}}{2}\phi_4 - \frac{1}{2}\phi_5 - \frac{\sqrt{3}}{2}\phi_6\right), \tag{9.55}$$

$$\psi_y = -\frac{2}{3}\left(\phi_2 - \frac{\sqrt{3}}{2}\phi_3 - \frac{1}{2}\phi_4 + \frac{\sqrt{3}}{2}\phi_5 - \frac{1}{2}\phi_6\right), \tag{9.56}$$

which correspond to relative translations of the two hexagonal lattices that make up the quasipattern These are slightly different from the definitions of the phase and phason modes used by Echebarria and Riecke, but they are equivalent, and are chosen such that the translation $x \to x + x_0$ implies $\phi_x \to \phi_x - x_0$ and so on To work out the phason modes, the translation is applied to just one of the two component lattices This leads to a difference in sign between ψ_y here and ψ_2 of Echebarria and Riecke (2001).

It can be shown that to linear order in the long-wave approximation the phase and phason modes satisfy the equations

$$\frac{\partial \phi}{\partial T} = D_1 \nabla^2 \phi + (D_2 - D_1)\nabla(\nabla \quad \phi), \tag{9.57}$$

$$\frac{\partial \psi_x}{\partial T} = D_3 \nabla^2 \psi_x + (D_4 - D_3)\frac{\partial}{\partial Y}\left(\frac{\partial \psi_x}{\partial Y} - \frac{\partial \psi_y}{\partial X}\right), \tag{9.58}$$

$$\frac{\partial \psi_y}{\partial T} = D_3 \nabla^2 \psi_y - (D_4 - D_3)\frac{\partial}{\partial X}\left(\frac{\partial \psi_x}{\partial X} - \frac{\partial \psi_y}{\partial Y}\right), \tag{9.59}$$

where $\phi = (\phi_x, \phi_y)$ is the phase field, and where

$$D_1 = \frac{1}{4} - \frac{q^2}{u_1}, \tag{9.60}$$

$$D_2 = \frac{3}{4} - \frac{2q^2}{v_1} - \frac{q^2}{u_1}, \tag{9.61}$$

$$D_3 = \frac{1}{4} - \frac{q^2}{u_2}, \tag{9.62}$$

$$D_4 = \frac{3}{4} - \frac{2q^2}{v_2} - \frac{q^2}{u_2}, \tag{9.63}$$

with

$$u_1 = 2R_0[\alpha + R_0(\beta + \gamma - \nu - \delta)], \tag{9.64}$$

$$u_2 = 2R_0[\alpha + R_0(\beta + \nu - \gamma - \delta)], \tag{9.65}$$

$$v_1 = -R_0[\alpha - 2R_0(\nu + 2\delta + 2\gamma + \beta)], \tag{9.66}$$

$$v_2 = -R_0[\alpha + 2R_0(\nu + 2\gamma - 2\delta - \beta)]. \tag{9.67}$$

The fact that the phase and phason modes decouple is a special feature of the dodecagonal case, and does not hold for octagonal or decagonal quasipatterns In the case $\gamma = \nu = 0$, when the two constituent hexagonal lattices decouple, the phase-diffusion coefficients, D_1 and D_2, take the same values as for a single hexagonal lattice (see equations (9.20) and (9.21)).

Substituting $\phi = \widehat{\phi} e^{\sigma T + ikX + ilY}$, $\psi_x = \widehat{\psi}_x e^{\sigma T + ikX + ilY}$ and $\psi_y = \widehat{\psi}_y e^{\sigma T + ikX + ilY}$, where $\widehat{\phi}$ is a constant vector, and $\widehat{\psi}_x$ and $\widehat{\psi}_y$ are constants, into equations (9 57)–(9.59) reveals four growth rate eigenvalues, namely

$$\sigma_1 = -D_1(k^2 + l^2), \tag{9.68}$$

$$\sigma_2 = -D_2(k^2 + l^2), \tag{9 69}$$

$$\sigma_3 = -D_3(k^2 + l^2), \tag{9 70}$$

$$\sigma_4 = -D_4(k^2 + l^2). \tag{9.71}$$

An instability occurs when any one of the diffusion coefficients changes sign from positive to negative The instabilities associated with D_1 and D_2 changing sign are the usual divergence-free $\nabla \cdot \phi = 0$ and irrotational $\nabla \times \phi = 0$ phase modes respectively The phason modes associated with D_3 and D_4 have eigenvectors that satisfy

$$\frac{\partial \psi_x}{\partial Y} - \frac{\partial \psi_y}{\partial X} = 0, \tag{9 72}$$

$$\frac{\partial \psi_x}{\partial X} + \frac{\partial \psi_y}{\partial Y} = 0, \tag{9 73}$$

respectively In large regions of parameter space, the band of stable dodecagonal quasipatterns is limited by the phason instability boundaries Examples of the stability regions are given in Figure 9.4 Echebarria and Riecke (2001) simulated the evolution of the phason instabilities numerically using equations (9.47) and (9 48), and found that both are subcritical, creating defects in the pattern The mode corresponding to $\sigma_3 > 0$ leads to a distorted quasipattern composed of Fourier modes of different wavenumbers (Figure 9 5). The simulation of the instability corresponding to $\sigma_4 > 0$ uses parameter values at which a square pattern is preferred, and patches of squares are seen to nucleate around the defects created by the instability and eventually to invade the whole cell (Figure 9.6) This is reminiscent of the triggering of the hexagon-roll transition by the Eckhaus instabilities of hexagons, discussed in the previous section.

9.2 Drift instabilities

The 2 : 1 mode interaction problem discussed in Section 6.2 exhibits a drift instability where travelling waves bifurcate from a stationary pattern Drift instabilities

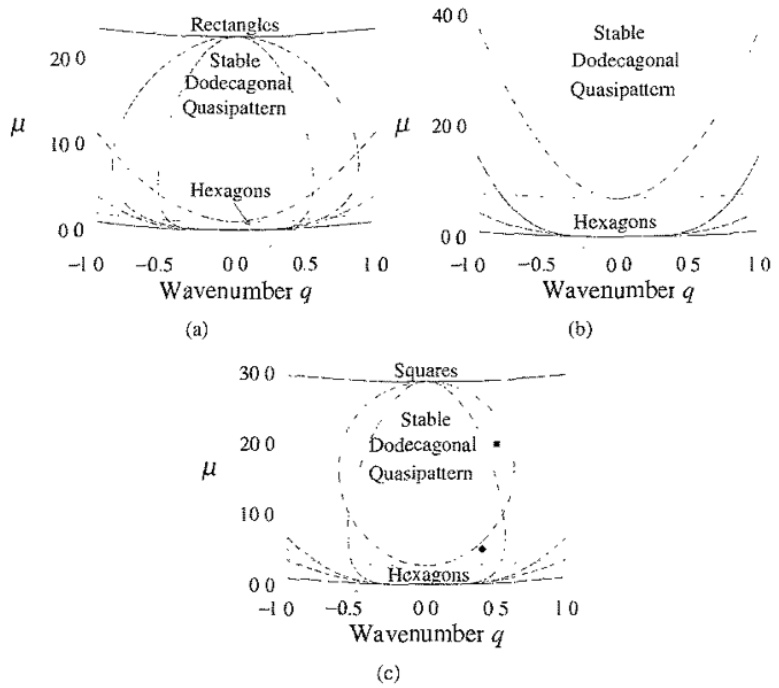


Fig. 9.4. Stability diagrams for dodecagonal quasipatterns with $\alpha = \beta = 1$ and (a) $\delta = 0\,7, \gamma = 0.3$ and $\nu = 1$, (b) $\delta = 0.9, \gamma = 0.8$ and $\nu = 0.8$ and (c) $\delta = 0.9$, $\gamma = 0.9$ and $\nu = 0\,4$ in equations (9 47) and (9 48) The lower solid line is the neutral stability curve, below which the quasipattern does not exist In all these cases, even a quasipattern at the preferred wavelength ($q = 0$) is unstable at onset, while hexagons are stable Hexagons lose stability to the quasipattern above the dotted line In cases (a) and (c) the quasipattern then loses stability to rectangles or squares respectively at the upper solid line The phase and phason instability boundaries are marked by dashed and dotted–dashed lines respectively. In cases (b) and (c) the region of stability is limited by the phason modes exclusively The diamond and square in (c) mark the initial conditions for the simulations in Figures 9 5 and 9.6 respectively. Reproduced from Echebarria, B and Riecke, H (2001) Sideband instabilities and defects of quasipatterns *Physica D*, **158**, 45–68, ©(2001), with permission from Elsevier.

are in fact a more general feature of patterns where there is coupling between a reflection-symmetry-breaking amplitude mode and the neutral phase mode associated with translation invariance A minimal, but general, way to investigate this is to write the solution as

$$u(x, t) = u_0(x + \phi(X, T)) + V(X, T)u_1(x), \tag{9 74}$$

where $u_0(x)$ is the basic reflection-symmetric pattern, $\phi(X, T)$ is its slowly varying phase, $u_1(x)$ is a reflection-symmetry-breaking eigenmode and $V(X, T)$ its slowly varying amplitude (Fauve, 1998). For example, we might have $u_0(x) \propto \cos(x)$ and $u_1(x) \propto \sin(x)$

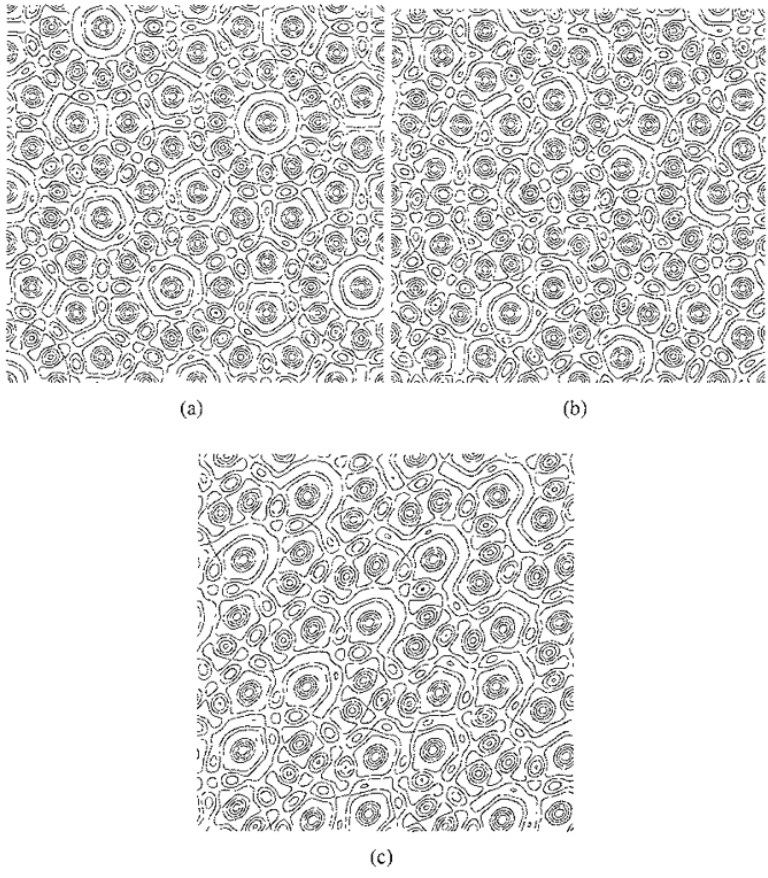


Fig. 9.5. Simulation of the instability corresponding to $\sigma_3 > 0$, where σ_3 is defined in equation (9.70). The parameters in equations (9.47) and (9.48) are $\mu = 5, \alpha = \beta = 1, \delta = 0.9, \gamma = 0.9$ and $\nu = 0.4$. The initial wavelength is given by $q = 0.4$, corresponding to the diamond in Figure 9.4c. The instability is seeded by adding a perturbation of the form $(\psi_x, \psi_y) \propto (1, 1)e^{i4\pi(X+Y)/L}$, where $L = 50$ is the length of each side of the simulation domain, thus satisfying equation (9.72). The pattern is shown at times (a) $T = 0$, (b) $T = 380$ and (c) $T = 450$. Reproduced from *Physica D*, **158**, Echebarria, B. and Riecke, H. (2001) Sideband instabilities and defects of quasipatterns 45–68, © (2001), with permission from Elsevier.

Now we have $u_1(-x) = -u_1(x)$, and so under the reflection symmetry $x \to -x$ the transformation $V \to -V$ is induced, along with the standard transformation $X \to -X$, $\phi \to -\phi$. There is also (normal form) symmetry under translations in x, $\phi \to \phi - x_0$, where x_0 is constant, and symmetry under translations in X, $X \to X - \Phi$, where Φ is constant, and in T, $T \to T - \tau$, where τ is constant. The evolution equations for the amplitude, V, and phase, ϕ, must be equivariant under these symmetries. Using appropriate scalings, which turn out to be $T = \epsilon^2 t$,

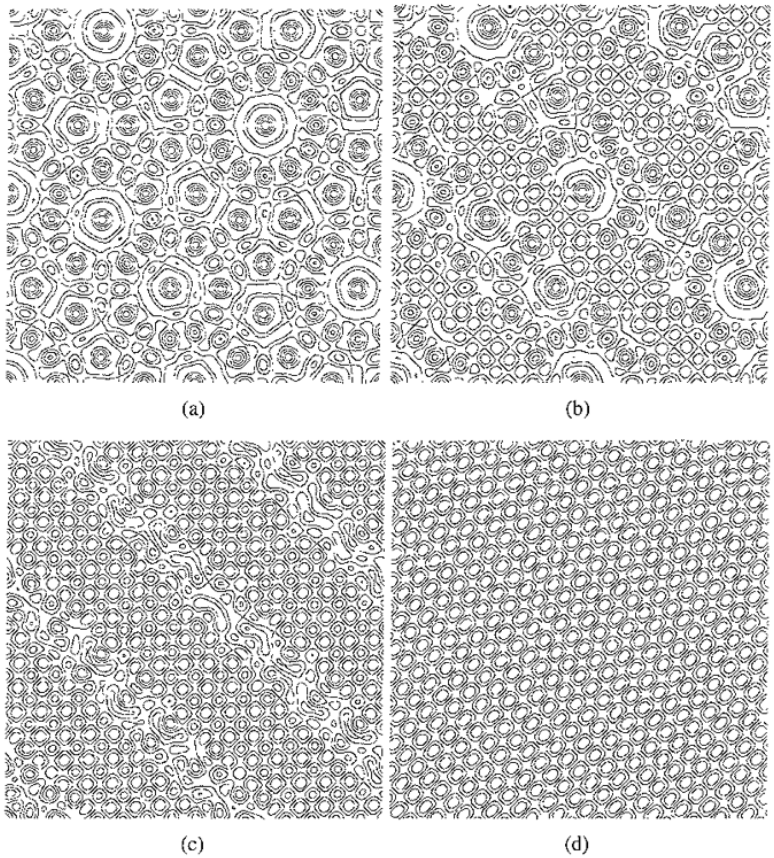


Fig 9.6. Simulation of the instability corresponding to $\sigma_4 > 0$, where σ_4 is defined in equation (9.71). The parameters in equations (9.47) and (9.48) are $\mu = 20$, $\alpha = \beta = 1$, $\delta = 0.9$, $\gamma = 0.9$ and $\nu = 0.4$. The initial wavelength is given by $q = 0.5$, corresponding to the square in Figure 9.4c. The instability is seeded by adding a perturbation of the form $(\psi_x, \psi_y) \propto (1, -1)e^{i4\pi(X+Y)/L}$, where $L = 25$ is the length of each side of the simulation domain, thus satisfying equation (9.73). The pattern is shown at times (a) $T = 0$, (b) $T = 440$, (c) $T = 445$ and (d) $T = 600$. Reproduced from Echebarria, B. and Riecke, H. (2001) Sideband instabilities and defects of quasipatterns, *Physica D*, **158**, 45–68, © (2001), with permission from Elsevier.

$X = \epsilon x$ and $V = \epsilon \widetilde{V}$, with ϕ and $\widetilde{V}$ being $O(1)$, the equations take the form

$$\epsilon^2 \frac{\partial \phi}{\partial T} = \epsilon a \widetilde{V} + \epsilon^2 b \frac{\partial^2 \phi}{\partial X^2} + \epsilon^2 c \widetilde{V} \frac{\partial \phi}{\partial X} + O(\epsilon^3), \tag{9.75}$$

$$\epsilon^3 \frac{\partial \widetilde{V}}{\partial T} = \epsilon \mu \widetilde{V} - \epsilon^3 \widetilde{V}^3 + \epsilon^2 d \frac{\partial^2 \phi}{\partial X^2} + \epsilon^3 f \frac{\partial^2 \widetilde{V}}{\partial X^2} + \epsilon^2 g \widetilde{V} \frac{\partial \phi}{\partial X} + \epsilon^3 h \widetilde{V} \frac{\partial \widetilde{V}}{\partial X} + \epsilon^3 s \frac{\partial \phi}{\partial X} \frac{\partial^2 \phi}{\partial X^2} + O(\epsilon^4) \tag{9.76}$$

To ensure a consistent scaling the constants a, d and g must be $O(\epsilon)$ and the constants b, c, f, h and s must be $O(1)$, while the bifurcation parameter μ is $O(\epsilon^2)$. The coefficient of $\widetilde{V}^3$ has been scaled to -1. If we set $a = \epsilon\widetilde{a}$, $d = \epsilon\widetilde{d}$, $h = \epsilon\widetilde{h}$ and $\mu = \epsilon^2\widetilde{\mu}$ and immediately drop all the tildes, including the one on $\widetilde{V}$, we get the following equations

$$\frac{\partial\phi}{\partial T} = aV + b\frac{\partial^2\phi}{\partial X^2} + cV\frac{\partial\phi}{\partial X}, \tag{9.77}$$

$$\frac{\partial V}{\partial T} = \mu V - V^3 + d\frac{\partial^2\phi}{\partial X^2} + f\frac{\partial^2 V}{\partial X^2} + gV\frac{\partial\phi}{\partial X} + hV\frac{\partial V}{\partial X} + s\frac{\partial\phi}{\partial X}\frac{\partial^2\phi}{\partial X^2} \tag{9.78}$$

to leading order, where now all the coefficients are $O(1)$. These equations are slightly modified from those found in Fauve, Douady and Thual (1991).

The stationary solution $V = 0$, $\phi = \phi_0$, where ϕ_0 is constant, loses stability when $\mu > 0$. The growth rate eigenvalue, σ, of perturbations proportional to e^{ikX} satisfies

$$\sigma^2 + \sigma\{-\mu + (b + f)k^2\} + (ad - b\mu)k^2 + bfk^4 = 0. \tag{9.79}$$

Under the assumption that perturbations are on long scales, so that $|k| \ll 1$, we can approximate $\sigma \approx \mu$, and so the bifurcation occurs at $\mu = 0$, as stated. The amplitude perturbation V now grows, and so the phase becomes time-dependent, according to equation (9.77). In other words, the pattern starts to drift.

The solution $V = V_0 \equiv \pm\sqrt{\mu}$, $\phi = aV_0 I$, representing a pattern drifting at a constant rate, bifurcates supercritically from the stationary solution. However the term $gV\partial\phi/\partial X$ destabilises the homogeneous drifting solution to spatially varying perturbations of the form $e^{\sigma T + ikX}$. The growth rate eigenvalues of the perturbation are

$$\sigma_1 = -2\mu + O(k), \tag{9.80}$$

$$\sigma_2 = \frac{ikV_0 ag}{2\mu} + k^2\left(\frac{V_0^2 ag(ag - h)}{8\mu^3} - b - \frac{ad}{2\mu}\right) + O(k^3) \tag{9.81}$$

For large enough ag, the real part of σ_2 is positive for $\mu > 0$, so the homogeneous solution is unstable. In this case localised travelling solutions of equations (9.77) and (9.78) are seen when the stationary pattern loses stability; further details can be found in Fauve (1991, 1998)

9.2.1 Drift instability of parametrically driven standing waves

Oscillatory patterns can also undergo drift instabilities. Consider a one-dimensional oscillatory pattern

$$u(x,t) = A(X,T)e^{i(t-x)} + B(X,T)e^{i(t+x)} + c.c.\,, \tag{9.82}$$

governed by the amplitude equations

$$\begin{aligned}\frac{\partial A}{\partial T} &= \mu A - c\frac{\partial A}{\partial X} + (1+i\alpha)\frac{\partial^2 A}{\partial X^2} - (\beta_r + i\beta_i)|A|^2 A \\ &\quad -(\gamma_r + i\gamma_i)|B|^2 A, \end{aligned} \tag{9.83}$$

$$\begin{aligned}\frac{\partial B}{\partial T} &= \mu B + c\frac{\partial B}{\partial X} + (1+i\alpha)\frac{\partial^2 B}{\partial X^2} - (\beta_r + i\beta_i)|B|^2 B \\ &\quad -(\gamma_r + i\gamma_i)|A|^2 B, \end{aligned} \tag{9.84}$$

where c, α, β_r, β_i, γ_r and γ_i are real constants, and where μ is a bifurcation parameter representing an external forcing that is constant in space and time. These equations are the same as equations (7.102) and (7.103), derived in Chapter 7 for counterpropagating waves in one spatial dimension, except that β_r is not set to 1 here. The scalings used in the derivation are $T = \epsilon^2 t$, $X = \epsilon x$, $\mu \sim \epsilon^2$, $c \sim \epsilon$, where $0 < \epsilon \ll 1$ with α, β, γ_r and γ_i all $O(1)$. Recall from Chapter 7 that these equations are only valid in the case of small group velocity, c.

Waves of the form (9.82) can also be generated by a time-periodic external forcing – usually there is an oscillating system parameter, and so this is known as **parametric forcing**. For example, in the Faraday wave experiment, described in Chapter 1, a horizontal layer of fluid is vibrated up and down, effectively modulating the acceleration due to gravity in the layer. This parametric forcing causes standing waves to appear on the fluid surface; typically these standing waves are subharmonic, oscillating with half the driving frequency. We shall assume here that we have an external forcing, $fe^{2it} + c.c.$, where f is a constant, so that the excited waves have unit frequency, as in equation (9.82). This forcing must enter the amplitude equations in a way that respects equivariance under the following symmetries:

(i) reflection in x: $X \to -X$, $A \leftrightarrow B$;
(ii) translation in x: $x \to x + \phi$, $A \to Ae^{i\phi}$, $B \to Be^{-i\phi}$, ϕ constant;
(iii) translation in X: $X \to X + \Phi$, $A(X,T) \to A(X-\Phi,T)$, $B(X,T) \to B(X-\Phi,T)$, Φ constant;
(iv) translation in t: $t \to t + \psi$, $A \to Ae^{-i\psi}$, $B \to Be^{-i\psi}$, $f \to fe^{-2i\psi}$, ψ constant;
(v) translation in T: $T \to T + \Psi$, $A(X,T) \to A(X,T-\Psi)$, $B(X,T) \to B(X,T-\Psi)$, Ψ constant

Only the translation in t transforms the parametric forcing. The leading order terms arising from the forcing are $f\bar{B}$ and $f\bar{A}$ in the $\partial A/\partial T$ and $\partial B/\partial T$ equations respectively. The imaginary part of f can be set to zero by choosing the origin of time appropriately. If we assume that $f \sim \epsilon^2$ then no further forcing terms arise to cubic order in the amplitude equations. We also allow for the standing wave frequency to be slightly different from half the driving frequency by including terms $i\omega A$ and $i\omega B$, where ω is a real constant representing the **detuning**. Since the frequency of the oscillation is determined by the driving frequency, and the wavenumber k is related to the frequency by a dispersion relation, this wavenumber is also fixed by the parametric forcing. We scale it to unity so that the solution can be written in the form (9.82). In the absence of the parametric forcing the real part of the dispersion relation had to take the form

$$\sigma = \mu - (1 - k^2)^2 + O([1 - k^2]^3), \tag{9.85}$$

so that the most unstable wavenumber would be $k_c = 1$ and the zero solution would be unstable for $\mu > 0$. Now the wavenumber is fixed by the forcing, so a term linear in k can appear on the righthand side of equation (9.85), which is equivalent to allowing an imaginary part to c. From now on we shall write $c = c_r + ic_i$, where c_r and c_i are real. Furthermore since the pattern is forced parametrically, μ need no longer force the growth of waves: in fact waves will appear even when μ is negative as we shall see. The leading order amplitude equations are now

$$\begin{aligned}\frac{\partial A}{\partial T} &= (\mu + i\omega)A + f\bar{B} - (c_r + ic_i)\frac{\partial A}{\partial X} + (1 + i\alpha)\frac{\partial^2 A}{\partial X^2} \\ &\quad -(\beta_r + i\beta_i)|A|^2 A - (\gamma_r + i\gamma_i)|B|^2 A, \end{aligned} \tag{9.86}$$

$$\begin{aligned}\frac{\partial B}{\partial T} &= (\mu + i\omega)B + f\bar{A} + (c_r + ic_i)\frac{\partial B}{\partial X} + (1 + i\alpha)\frac{\partial^2 B}{\partial X^2} \\ &\quad -(\beta_r + i\beta_i)|B|^2 B - (\gamma_r + i\gamma_i)|A|^2 B, \end{aligned} \tag{9.87}$$

where the coefficient of the parametric forcing terms is set to one by scaling f. If f is nonzero then there can be no solutions where one of A or B is zero and the other is nonzero, because the parametric forcing term would cause the zero amplitude to grow: left-travelling waves force the growth of right-travelling waves and vice versa. This means we do not expect to see simple travelling waves. Parametrically driven standing waves are possible, however, and they are subject to a drift instability at high values of the forcing, f. We will use the method of Fauve, Douady and Thual (1991) to analyse this situation.

Equations (9.86) and (9.87) are quite complicated so we will only consider spatially homogeneous solutions in what follows. Linearising around the zero solution

gives

$$\frac{dA}{dT} = (\mu + i\omega)A + f\bar{B}, \tag{9.88}$$

$$\frac{dB}{dT} = (\mu + i\omega)B + f\bar{A}. \tag{9.89}$$

Substituting $A = \hat{A}e^{\sigma T}$ and $B = \hat{B}e^{\sigma T}$, where $\hat{A}$ and $\hat{B}$ are complex constants, gives the growth rate eigenvalues

$$\sigma = \mu \pm \sqrt{f^2 - \omega^2}, \tag{9.90}$$

so the zero solution is always unstable if $\mu > 0$ and is unstable for $\mu < 0$ if $f^2 > \mu^2 + \omega^2$. To investigate nonzero solutions we set $A = e^{R+S+i\theta+i\phi}$ and $B = e^{R-S+i\theta-i\phi}$, giving

$$\frac{dR}{dT} = \mu + \cosh 2S(f\cos 2\theta - (\beta_1 + \gamma_1)e^{2R}), \tag{9.91}$$

$$\frac{dS}{dT} = -\sinh 2S(f\cos 2\theta + (\beta_r - \gamma_r)e^{2R}), \tag{9.92}$$

$$\frac{d\theta}{dT} = \omega - \cosh 2S(f\sin 2\theta + (\beta_i + \gamma_i)e^{2R}), \tag{9.93}$$

$$\frac{d\phi}{dT} = \sinh 2S(f\sin 2\theta - (\beta_i - \gamma_i)e^{2R}) \tag{9.94}$$

The equation for ϕ decouples from the others: ϕ is the spatial phase associated with a translation of the origin in x, and so gives rise to a neutral mode. A standing wave solution, where $dA/dT = dB/dT = 0$ and $|A| = |B|$ is given by $S = 0$, with ϕ any real constant, and R and θ satisfying

$$\mu + f\cos 2\theta - (\beta_r + \gamma_r)e^{2R} = 0, \tag{9.95}$$

$$\omega - f\sin 2\theta - (\beta_i + \gamma_i)e^{2R} = 0 \tag{9.96}$$

Eliminating θ between these two equations gives

$$e^{4R}[(\beta_1 + \gamma_r)^2 + (\beta_i + \gamma_i)^2] + e^{2R}[\omega(\beta_r + \gamma_r) + \mu(\beta_i + \gamma_i)] + \mu^2 + \omega^2 - f^2 = 0, \tag{9.97}$$

which has one positive solution for e^{2R} in the regime

$$\mu < 0 \tag{9.98}$$

$$f^2 > \mu^2 + \omega^2 \tag{9.99}$$

that we shall consider from now on. There are corresponding solutions $R = R_0$ and $\theta = \theta_0$. Perturbing around this solution such that $R = R_0 + \hat{R}$, $S = \hat{S}$, $\theta = \theta_0 + \hat{\theta}$

and $\phi = \phi_0 + \widehat{\phi}$, where $|\widehat{R}|, |\widehat{S}| \ll R_0$, $|\widehat{\theta}| \ll \theta_0$, $|\widehat{\phi}| \ll \phi_0$, we find

$$\frac{\mathrm{d}\widehat{R}}{\mathrm{d}T} = -2(\mu + f\cos 2\theta_0)\widehat{R} - 2f\widehat{\theta}\sin 2\theta_0, \tag{9.100}$$

$$\frac{\mathrm{d}\widehat{\theta}}{\mathrm{d}T} = -2(\omega - f\sin 2\theta_0)\widehat{R} - 2f\widehat{\theta}\cos 2\theta_0, \tag{9.101}$$

$$\frac{\mathrm{d}\widehat{S}}{\mathrm{d}T} = 2(\mu - 2\beta_r e^{2R_0})\widehat{S}, \tag{9.102}$$

$$\frac{\mathrm{d}\widehat{\phi}}{\mathrm{d}T} = 2(\omega - 2\beta_i e^{2R_0})\widehat{S} \tag{9.103}$$

The growth rate eigenvalues, σ, corresponding to the perturbations in R and θ satisfy

$$\sigma^2 + 2\sigma(\mu + 2f\cos 2\theta_0) + 4f(f - \omega\sin 2\theta_0) = 0. \tag{9.104}$$

Both roots will be negative, and so perturbations in R and θ will be damped if

$$\mu + 2f\cos 2\theta_0 > 0, \tag{9.105}$$

$$f(f + \mu\cos 2\theta_0 - \omega\sin 2\theta_0) > 0 \tag{9.106}$$

hold. Using equations (9.95) and (9.96) these inequalities can be rewritten as

$$2e^{2R_0}(\beta_r + \gamma_r) - \mu > 0, \tag{9.107}$$

$$e^{4R_0}[(\beta_r + \gamma_r)^2 + (\beta_i + \gamma_i)^2] - e^{2R_0}[\mu(\beta_r + \gamma_r) + \omega(\beta_i + \gamma_i)] > 0 \tag{9.108}$$

We have assumed that $\mu < 0$, so $\beta_r + \gamma_r > 0$ is a sufficient condition for the first inequality to hold. Then the second will be satisfied if the detuning, ω, is small enough – a sufficient condition is $|\omega| < |\mu(\beta_r + \gamma_r)/(\beta_i + \gamma_i)|$. Looking at perturbations in S and ϕ now, we see that close to onset, where the standing wave amplitude, e^{2R_0}, is small, perturbations in S, and hence in ϕ, will be damped in the regime $\mu < 0$, and the standing wave solution will be stable. As the driving, f, increases, so does the amplitude e^{2R_0}, since eliminating θ between equations (9.95) and (9.96) we find

$$f^2 = \mu^2 + \omega^2 - 2[\mu(\beta_r + \gamma_r) + \omega(\beta_i + \gamma_i)]e^{2R_0} + [(\beta_r + \gamma_r)^2 + (\beta_i + \gamma_i)^2]e^{4R_0}, \tag{9.109}$$

where the coefficient of e^{2R_0} is positive under the assumptions we have made. If $\beta_i < 0$, then as the amplitude, e^{2R_0}, grows equation (9.102) shows that eventually the growth rate of perturbations in S will become positive, and so $|S|$ will grow from zero, breaking the x-reflection symmetry; this causes the spatial phase, ϕ, to grow, and so the pattern starts to drift

9.3 Galilean invariance and flat modes

In Section 8.6 we saw that Galilean symmetry plays a role in the skew-varicose and oscillatory instabilities of rolls. We will now look in more detail at systems that are Galilean-invariant, in other words invariant under the transformation to a frame moving at constant velocity. When the quantity, $u(x,t)$, in the governing partial differential equation

$$\frac{\partial u(x,t)}{\partial t} = f(u(x,t),\mu), \tag{9.110}$$

is itself a component of velocity, Galilean-invariance is sometimes natural. This is the case, for example, in Rayleigh–Bénard convection with stress-free boundary conditions (see, for example, Coullet and Fauve, 1985). This situation has been studied by Matthews and Cox (2000a), among others, and we will mostly follow their approach here.

In one space dimension, with $x \in \mathbb{R}$, the variables u and x transform as

$$x \to x - vt, \tag{9.111}$$

$$u \to u + v, \tag{9.112}$$

where v is the constant speed of translation. Introducing the transformed variables

$$x' = x - vt, \tag{9.113}$$

$$t' = t \tag{9.114}$$

we find that

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial x'}, \tag{9.115}$$

$$\frac{\partial}{\partial t} = -v\frac{\partial}{\partial x'} + \frac{\partial}{\partial t'}, \tag{9.116}$$

and so the Galilean transformation can be summarised as

$$\left\{u \to u + v,\ \frac{\partial}{\partial t} \to \frac{\partial}{\partial t} - v\frac{\partial}{\partial x}\right\} \tag{9.117}$$

It is straightforward to check that the terms $\partial u/\partial t$ and u are not Galilean-invariant, but $(\partial u/\partial t + u\partial u/\partial x)$ and $\partial u/\partial x$ are, so a Galilean-invariant governing equation must take the form

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} + g\left(\frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2}, \ldots\right), \tag{9.118}$$

where $g(\partial u/\partial x, \partial^2 u/\partial x^2, \ldots)$ contains only x-derivatives of u and their products. If the system also has x-reflection symmetry,

$$x \to -x, \tag{9.119}$$

$$u \to -u, \tag{9.120}$$

then the leading order linear term is the second derivative, and so we have

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2} - u\frac{\partial u}{\partial x} + \cdots, \tag{9.121}$$

where D is a positive constant diffusion coefficient.

Linearising around the zero solution, $u = 0$, the growth rate of long-wave disturbances is found to be

$$\sigma = -Dk^2 + O(k^4) \tag{9.122}$$

The zero wavenumber, **flat** or **Goldstone**, mode is neutrally stable, and long-wave modes are very nearly so. This means we can't neglect them even if modes at nonzero wavenumber are growing, as is the case when the linear part of the governing equation is

$$\frac{\partial u}{\partial t} = -\frac{\partial^2}{\partial x^2}\left[\mu u - \left(1 + \frac{\partial^2}{\partial x^2}\right)^2 u\right], \tag{9.123}$$

where μ is a constant, and the dispersion relation for disturbances to the zero solution is

$$\sigma = k^2[\mu - (1 - k^2)^2]. \tag{9.124}$$

When $\mu = 0$, modes $k^2 = 1$ are neutral (as is the $k = 0$ mode), and when $\mu = \hat{\mu}\epsilon^2$ is nonzero, with $\hat{\mu}$ being $O(1)$ and $|\epsilon| \ll 1$, a small band of wavenumbers around $k^2 = 1$ will grow as shown in Figure 9.7. Although there is a stationary bifurcation at $\mu = 0$ for modes with $k^2 = 1$, the Ginzburg–Landau equation does not describe this situation because the $k = 0$ mode is not damped.

To study a situation where both a spatially uniform, or flat, mode and a one-dimensional finite-wavenumber mode contribute to the pattern we write

$$u(x, t) = A(X, T)e^{ix} + \bar{A}(X, T)e^{-ix} + B(X, T) + h.o.t., \tag{9.125}$$

where the dependence of the amplitudes on the long scales X and T allows for modulation. The observable, $u(x, t)$, must be real, so although $A(X, T)$ can be complex, $B(X, T)$ has to be real. We can write down amplitude equations for A and B using equivariance conditions in the usual way. For a Galilean-invariant Euclidean-symmetric system, the symmetries that must be respected are reflection

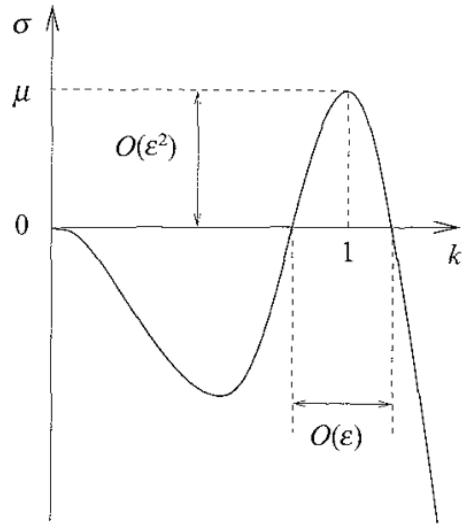


Fig. 9.7. The growth rate, σ, of Fourier mode solutions, $e^{\sigma t+ikx}$, to equation (9.123) for small $\mu \sim \epsilon^2$, where $|\epsilon| \ll 1$. The growth rate is positive for wavenumbers in a band of size $O(\epsilon)$ around $k = 1$, and zero for the flat mode $k = 0$.

in x, given by

$$X \to -X, \; A \to -\bar{A}, \; B \to -B, \tag{9.126}$$

and translation into a frame moving with constant velocity, given by

$$x \to x + vT, \; A \to Ae^{-ivT}, \; B \to B + v, \tag{9.127}$$

where v is constant, along with the usual symmetries under translation in x,

$$x \to x + x_0, \; A \to Ae^{-ix_0}, \; B \to B, \tag{9.128}$$

where x_0 is a constant, translation in X

$$X \to X + X_0, \; A(X, T) \to A(X - X_0, T), \; B(X, T) \to B(X - X_0, T), \tag{9.129}$$

where X_0 is a constant, and translation in T,

$$T \to T + T_0, \; A(X, T) \to A(X, T - T_0), \; B(X, T) \to B(X, T - T_0), \tag{9.130}$$

where T_0 is a constant. The leading order amplitude equations equivariant under these symmetries are

$$\frac{\partial A}{\partial T} = \mu A + \frac{\partial^2 A}{\partial X^2} - iAB, \tag{9.131}$$

$$\frac{\partial B}{\partial T} = \nu \frac{\partial^2 B}{\partial X^2} - \frac{\partial}{\partial X}|A|^2, \tag{9.132}$$

where all coefficients but $\mu \in \mathbb{R}$ and $\nu > 0$ ($\in \mathbb{R}$) have been set to ± 1 by scaling A and X. The sign of the diffusion terms is chosen so that short wavelength modes decay. The sign of the linear growth term A can be either positive or negative, but we choose here to look at the situation where it is positive and the zero solution is linearly unstable, so that we can investigate pattern formation. The coefficient of iAB in equation (9.131) is set by the Galilean symmetry (9.127), and cannot be zero. The sign of the term $-\partial|A|^2/\partial X$ in equation (9.132) can be changed by setting $\widehat{X} = -X$ and immediately dropping the hat: this means we need only consider one choice of sign here. The coefficient of this term will, however, only be nonzero if the governing partial differential equation for u has quadratic nonlinear terms, so this coupling term is not guaranteed to appear. Consistent scalings are μ, $\partial/\partial T \sim \epsilon^2$, $\partial/\partial X \sim \epsilon$, $A \sim \epsilon^{3/2}$ and $B \sim \epsilon^2$, with ν being $O(1)$.

Since $u(x, t)$ is a velocity, the amplitude B represents a large-scale flow that is generated by spatial inhomogeneities in the amplitude, A, of the periodic pattern, which in turn is advected by the large-scale flow. If we write $A = Re^{i\phi}$ then the coupling term $-iAB$ contributes to the phase equation according to

$$\frac{\partial \phi}{\partial T} = \cdot \quad - B, \tag{9.133}$$

so the advective flow leads to a drift in the phase of the periodic pattern as seen in the fixed frame. This is to be expected since the phase seen at a fixed point of reference will vary as the pattern moves past. The coupling term does not affect the evolution of R.

A cubic term, $|A|^2A$, appears at higher order in the equation for $\partial A/\partial T$. If its coefficient is negative it is stabilising, but since it does not occur at leading order it cannot quench the linear instability given by the μA term for $\mu > 0$, so there are no stationary stripe solutions. Instead equations (9.131) and (9.132) have exponentially growing stripe solutions of the form

$$A = A_0 e^{\mu T}, \quad B = 0, \tag{9.134}$$

where A_0 is a constant. Perturbing this solution by setting $A = A_0 e^{\mu T}(1 + a_r + ia_i)$ and $B = b$, where $|a_r|, |a_i|, |b| \ll 1$, and linearising we find

$$\frac{\partial a_r}{\partial T} = \frac{\partial^2 a_i}{\partial X^2}, \tag{9.135}$$

$$\frac{\partial a_i}{\partial T} = \frac{\partial^2 a_i}{\partial X^2} - b, \tag{9.136}$$

$$\frac{\partial b}{\partial T} = \nu \frac{\partial^2 b}{\partial X^2} - 2A_0^2 e^{2\mu T} \frac{\partial a_r}{\partial X} \tag{9.137}$$

If the perturbations have no space dependence, then both a_r and b are neutral modes, arising from the translation and Galilean symmetries respectively For a mode of wavenumber k, a_r satisfies

$$\frac{\partial a_r}{\partial T} = -k^2 a_r, \tag{9.138}$$

and so $a_r = a_0 e^{-k^2 T}$, where a_0 is a constant. As a result $\partial b/\partial T$, and hence b, has a component proportional to $e^{(2\mu - k^2)T}$ that grows faster than the basic stripe solution for long-wave (small k) perturbations. We can conclude that the exponentially growing stripe solution is unstable to modes with spatially modulated amplitudes. In fact numerical simulations of the amplitude equations (9.131) and (9.132) (Matthews & Cox, 2000a) show that solutions do not grow exponentially, but remain bounded.

9.4 Conservative systems and flat modes

Flat modes are also important in systems with a conservation law of the form

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x} F(u), \tag{9.139}$$

where F is a nonlinear operator that can involve spatial derivatives. The marker quantity, u, is conserved in the sense that

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathcal{D}} u \, \mathrm{d}x = 0 \tag{9.140}$$

holds under suitable boundary conditions, where $\mathcal{D}$ is the whole spatial domain We shall analyse conservative systems with a reflection symmetry, following Matthews and Cox (2000b)

There are two possibilities for the reflection symmetry: either the quantity u can remain unchanged under reflection, so that

$$x \to -x, \tag{9.141}$$

$$u \to u, \tag{9.142}$$

or it can change sign, so that

$$x \to -x, \tag{9.143}$$

$$u \to -u. \tag{9.144}$$

We shall first consider the case where u is unchanged, which is relevant to situations where u represents something like a density, rather than a velocity For example, in the Faraday wave problem the marker quantity is $h(x, y, t)$, the deviation

of the height of the free surface of the fluid from its position at rest, which must satisfy

$$\int_D h \, \mathrm{d}x = 0 \tag{9.145}$$

for an incompressible fluid, since the amount of fluid in the container is fixed. Clearly the height of the free surface is not changed under a horizontal reflection, so we have a reflection symmetry of the required type. There are many other examples of such systems, among them thermosolutal convection and magnetoconvection (Matthews & Cox, 2000b) and the formation of sand banks and sand waves offshore (Komarova & Newell, 2000).

In a system with a conservation law of the form (9.139) all the linear terms must be derivatives of u, and since we have the reflection symmetry, the leading order term must be a multiple of $\partial^2 u/\partial x^2$, so we will have a zero wavenumber mode as before (Figure 9.7). Assuming we also have a finite wavenumber mode we can once again write the pattern in the form

$$u(x,t) = A(X,T)e^{ix} + \bar{A}(X,T)e^{-ix} + B(X,T) + h.o.t \tag{9.146}$$

The reflection symmetry gives the transformation

$$x \to -x, \quad A \to \bar{A}, \quad B \to B. \tag{9.147}$$

The remaining symmetries are the translations given by equations (9.128)–(9.130). In addition we know that all terms in the $\partial B/\partial T$ equation must be spatial derivatives, since B is conserved over all space. Taking all this into consideration the new amplitude equations must be

$$\frac{\partial A}{\partial T} = \mu A + \frac{\partial^2 A}{\partial X^2} - |A|^2 A - AB, \tag{9.148}$$

$$\frac{\partial B}{\partial T} = \nu \frac{\partial^2 B}{\partial X^2} + \kappa \frac{\partial^2}{\partial X^2}|A|^2 \tag{9.149}$$

at leading order. The scalings used in this case are $\mu, \partial/\partial T \sim \epsilon^2$, $\partial/\partial X \sim \epsilon$, $A \sim \epsilon$ and $B \sim \epsilon^2$, where $0 < \epsilon \ll 1$, with ν and κ real $O(1)$ constants. Three of the coefficients have been scaled to ± 1 by scaling A, B and X. Again the sign of the diffusion terms is chosen to ensure that short-scale modes decay, so we must have $\nu > 0$. The sign of the term $-|A|^2 A$ has been chosen to ensure that stripes exist in $\mu > 0$ when $B = 0$. The sign of the term $-AB$ can be changed by setting $B = -\widehat{B}$ and $\kappa = -\widehat{\kappa}$ and immediately dropping the hats, so we need only consider one choice. Now both coupling terms will vanish if there are no quadratic nonlinear terms in the governing partial differential equation for u, and higher order coupling terms will need to be included.

The stripe solution of equations (9.148) and (9 149) is given by

$$A = (\mu - q^2)^{1/2} e^{iqX} \equiv A_0, \tag{9.150}$$

$$B = 0 \tag{9 151}$$

Perturbing it by setting

$$A = A_0(1 + a(T)e^{ikX} + \overline{c(T)}e^{-ikX}), \tag{9 152}$$

$$B = b(T)e^{ikX} + \overline{b(T)}e^{-ikX}, \tag{9.153}$$

where $|a|, |b|, |c| \ll 1$ and substituting into the amplitude equations gives

$$\frac{db}{dT} = -\nu k^2 b - \kappa |A_0|^2 k^2 (a + c), \tag{9.154}$$

$$\frac{da}{dT} = -|A_0|^2 (a + c) - k^2 a - 2qka - b, \tag{9 155}$$

$$\frac{dc}{dT} = -|A_0|^2 (a + c) - k^2 c + 2qkc - b. \tag{9.156}$$

If $a, b, c \propto e^{\sigma T}$, then the growth rate eigenvalues, σ, satisfy

$$\sigma^3 + \sigma^2[(2 + \nu)k^2 + 2|A_0|^2] + \sigma[(1 + \nu)k^4 - 4\nu q^2 k^2 + 2k^2|A_0|^2(1 + \nu - \kappa)] + k^4[\nu k^2 - 4\nu q^2 - 2|A_0|^2(\kappa - \nu)] = 0. \tag{9.157}$$

If the product of the roots of this equation is positive,

$$-k^4[\nu k^2 - 4\nu q^2 - 2|A_0|^2(\kappa - \nu)] > 0, \tag{9 158}$$

then at least one growth rate eigenvalue has positive real part, and the original stripe pattern is unstable Since ν is positive, the inequality (9.158) is most likely to hold for perturbation modes with $k^2 \to 0$, and the instability criterion becomes

$$4\nu q^2 + 2|A_0|^2(\kappa - \nu) > 0. \tag{9.159}$$

Stripes at all values of q are unstable if $\kappa > \nu$. This means that even patterns at the critical wavelength, which have $q = 0$ are unstable If $\kappa < \nu$ then small-amplitude stripes, with

$$|A_0|^2 < \frac{2\nu q^2}{\nu - \kappa}, \tag{9 160}$$

are unstable.

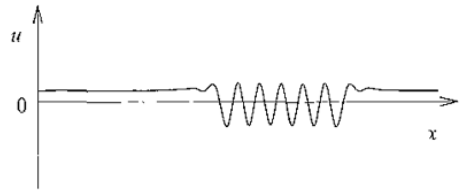


Fig 9 8 Sketch of a strongly modulated localised nonlinear pulse solution to equation (9 164).

For stripes at the critical wavenumber ($q = 0$) equations (9 154)–(9.156) can be rearranged to give

$$\frac{\mathrm{d}b}{\mathrm{d}T} = -\nu k^2 b - \kappa |A_0|^2 k^2 (a + c), \tag{9 161}$$

$$\frac{\mathrm{d}}{\mathrm{d}T}(a + c) = -(2|A_0|^2 + k^2)(a + c) - 2b, \tag{9.162}$$

$$\frac{\mathrm{d}}{\mathrm{d}T}(a - c) = -k^2 (a - c), \tag{9.163}$$

so the phase perturbation, $(a - c)$, is governed by a growth rate eigenvalue $-k^2$ and does not grow. The unstable mode in the case $\kappa > \nu$ arises through an interaction of the amplitude perturbation, $(a + c)$, and the zero-wavenumber mode, b. The mechanism of this instability for $q = 0$ is therefore different from the standard Eckhaus instability of stripes, which is a phase mode. Matthews and Cox (2000a) go on to show that the bifurcation is supercritical for $q = 0$, while Proctor (2001) points out that the inclusion of higher-order terms in the amplitude equations shows that in wide domains the bifurcation will be subcritical for arbitrary q, and that for $q \neq 0$ the standard Eckhaus phase diffusion equation can be derived using the Pomeau–Manneville (1979) phase approximation.

For the particular example of a system governed by the variant Swift–Hohenberg model

$$\frac{\partial u}{\partial t} = -\frac{\partial^2}{\partial x^2}\left[\mu u - \left(1 + \frac{\partial^2}{\partial x^2}\right)^2 u - s u^2 - u^3\right], \tag{9 164}$$

where μ is a real bifurcation parameter and s is a real constant, Matthews and Cox (2000a) find strongly modulated localised solutions in the nonlinear régime, such as the one shown in Figure 9 8. Norbury, Wei and Winter (2002) present existence and stability results for some classes of such solutions.

For conservative systems where u changes sign under reflection $\{x \to -x,\ u \to -u\}$ it turns out that the amplitude equations for A and B are the same as those derived for the case of Galilean symmetry, namely equations (9 131) and

(9.132) The reflection symmetry acts exactly as in equation (9.126) and the translation symmetries in x (9.128), X (9.129) and T (9.130) also apply. In the Galilean-invariant case a term proportional to B in the $\partial B/\partial T$ equation is forbidden by the Galilean symmetry (9.127), while in the conservative case it is forbidden because all terms in the $\partial B/\partial T$ equation must be spatial derivatives, since B is conserved over all space, as shown by equation (9.140). The amplitude equations take the same form in both cases, even though their derivation is slightly different. A further difference is that now the existence of the coupling term $-iAB$ in equation (9.131) depends on there being quadratic terms in the governing equation for u as it is no longer guaranteed by the Galilean invariance. If the term is present, then to set its coefficient to $-i$ we must scale B.

Exercises

9.1 Analyse the phase instabilities of a square pattern

$$u(x, y, t) = A(X, Y, T)e^{ix} + B(X, Y, T)e^{iy} + c.c. + h.o.t.,$$

where $A(X, Y, T) = R_0 e^{iqX}$ and $B(X, Y, T) = R_0 e^{iqY}$, for real constants R_0 and q such that the pattern is amplitude-stable, using methods similar to those shown for hexagons in Section 9.1.1.

(The full analysis can be found in Hoyle, 1993.)

9.2 Analyse the phase instabilities of oscillatory standing squares

$$\begin{aligned} u(x, y, t) = {} & A(X, Y, T)e^{i(t-x)} + B(X, Y, T)e^{i(t+x)} \\ & + C(X, Y, T)e^{i(t-y)} + D(X, Y, T)e^{i(t+y)} \\ & + c.c. + h.o.t., \end{aligned}$$

where $A = R_0 e^{i(\Omega T - qX)}$, $B = R_0 e^{i(\Omega T + qX)}$, $C = R_0 e^{i(\Omega T - qY)}$, $D = R_0 e^{i(\Omega T + qY)}$, for real constants R_0, q and Ω such that the pattern is amplitude-stable, using similar methods to those used in the previous exercise.

(The full analysis can be found in Hoyle, 1994b.)

9.3 Analyse the phase and phason instabilities of octagonal quasipatterns

$$u(x, y, t) = \sum_{n=1}^{4} A_n(X, Y, T)e^{ik_n \cdot x} + c.c. + h.o.t.,$$

where $k_n = (\cos\{(n-1)\pi/4\}, \sin\{(n-1)\pi/4\})$ and $A_n = R_0 e^{iqk_n \cdot X}$ for $n = 1, 2, 3, 4$ for real constants R_0 and q such that the pattern is amplitude-stable, using methods similar to those used for dodecagonal quasipatterns in Section 9.1.2.

If you are feeling particularly keen, do the same for decagonal quasipatterns (which have tenfold rotational symmetry).

(Both of these analyses can be found in Echebarria & Riecke, 2001.)

9.4 Starting from the conservative variant Swift–Hohenberg equation

$$\frac{\partial u}{\partial t} = -\frac{\partial^2}{\partial x^2}\left[\mu u - \left(1 + \frac{\partial^2}{\partial x^2}\right)^2 u - su^2 - u^3\right],$$

where μ is a real bifurcation parameter and s is a real constant, and writing

$$u(x,t) = (\epsilon A(X,T)e^{ix} + c.c.) + \epsilon^2 B(X,T) + (\epsilon^2 C(X,T)e^{2ix} + c.c.) + O(\epsilon^3),$$

derive the amplitude equations (9.148) and (9.149), identifying the coefficients ν and κ in terms of s.

What happens if $s = 0$? Will every system with a conservation law of the form (9.139) and a reflection symmetry $\{x \to -x,\ u \to u\}$ give rise to equations (9.148) and (9.149)?

(The answers to this exercise can be found in Matthews & Cox, 2000b)

10

Spirals, defects and spiral defect chaos

In real systems, patterns are rarely perfect: they usually contain some defects – places where the pattern is irregular in some way. This chapter will focus on the shape and movement of defects in natural patterns. Spirals are a special kind of defect, so common and regular that they are often thought of as patterns in their own right. In the form of spiral defect chaos they compete with rolls in large aspect ratio systems.

There are very many ways in which a pattern can be imperfect, and a multitude of interesting localised structures that can be seen in pattern-forming systems. We will only look at a few of the most common defects, such as spirals and the closely related target patterns, dislocations, where two stripes merge into one, and grain boundaries, which join patches of pattern at different orientations. For those who want to know more about defects and coherent structures, the review article by Cross and Hohenberg (1993) has further information, and more recently Sandstede and Scheel (2004) have attempted a systematic classification of defects in oscillatory media.

We will start with isolated defects in otherwise regular patterns close to onset.

10.1 Types of isolated defect

A roll pattern modulated on long space and time scales can be written in the form

$$u(x, y, t) = A(X, Y, T)e^{ix} + c.c. \tag{10.1}$$

at leading order, as discussed in Chapter 7, where X, Y, T are long scales in the x, y, t directions respectively. The function $A(X, Y, T)$ is the envelope, and the carrier wave is given by e^{ix}. At a defect in the pattern, the distinction between the envelope and the carrier wave breaks down because the short lengthscale behaviour is no longer just a simple wave. Alternatively you can say that the

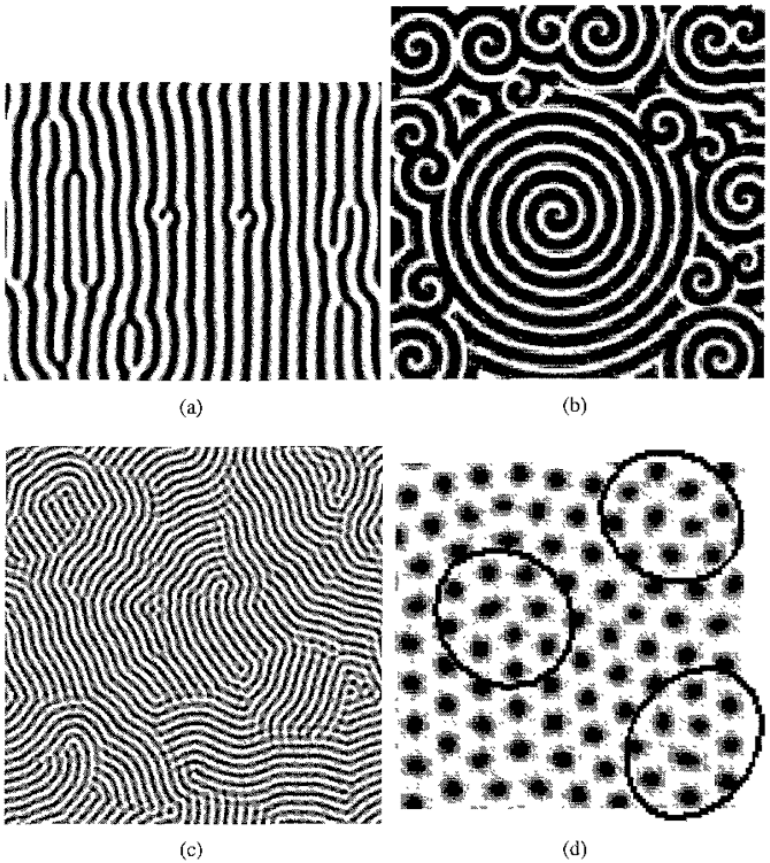


Fig 10.1. Examples of defects: (a) dislocations in a roll pattern during a numerical simulation of the Ginzburg–Landau equation, (b) spirals in a simulation of the FitzHugh–Nagumo equations, (c) a numerical simulation of a Swift–Hohenberg model, showing several grain boundaries between patches of rolls at different orientations, (d) penta-hepta defects (circled) of a hexagonal pattern in a numerical simulation of a Swift–Hohenberg model, (e) diagram of a focus singularity, (f) diagram of a convex disclination (several convex disclinations can also be seen in (c)), (g) a concave disclination and (h) a wall focus. Spiral picture courtesy of and ©Steve Tobias, University of Leeds, 2003

envelope varies on a short spatial scale, and of course the point of envelopes is that they should only vary on long scales and leave all the short scale variation in the carrier wave.

Defects can be associated with jumps in the phase, that is with singularities of the phase gradient. At the same time, the amplitude of the pattern must be zero so that $u(x, y, t)$ remains defined at every point. The dislocation defect of rolls, and the core of spirals in oscillatory or excitable systems shown in Figures 10.1a and b are examples of this kind of defect. There are also defects that are joins

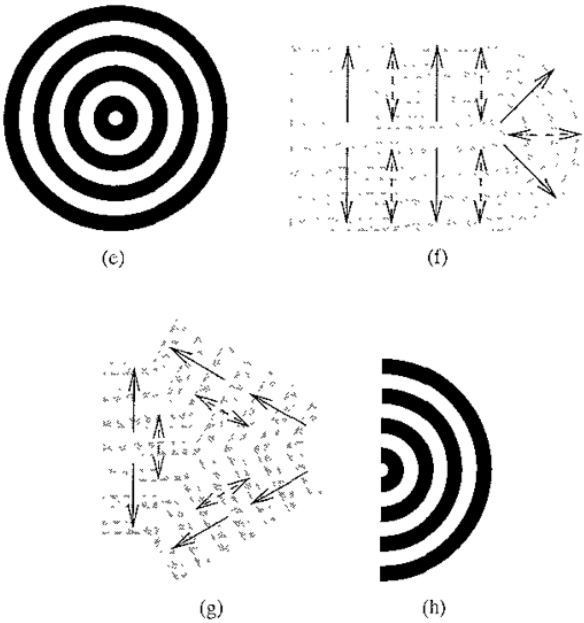


Fig 10 1 (*cont*)

between sections of pattern with different orientations, for example grain boundaries, illustrated in Figure 10 1c The penta-hepta pair in a hexagonal pattern is a more exotic example of a defect. The amplitudes of two of the three rolls making up the hexagonal pattern are zero at the heart of this defect, and instead of two hexagons, a pentagon and a heptagon are seen (Figure 10.1d).

Roll patterns in large domains with sides that are many roll wavelengths long, or in irregular geometries, tend to have many defects. Along with grain boundaries and dislocations, focus singularities, and convex and concave disclinations are common (Figure 10 1e–g). Wall foci (Figure 10.1h), which can be thought of either as partial focus singularities or convex disclinations at the wall, are often forced by the geometry in convection because it is energetically preferable for rolls to align perpendicular to the boundary (see Section 10 5 below)

10.2 Dislocation of a roll pattern

A dislocation is a defect of a roll pattern where two rolls merge into one, thus increasing the local wavelength. As usual, we write the roll pattern as the product of a modulated envelope and a carrier wave

$$u(x, y, t) = A(X, Y, T)e^{ix} + c\,c. \tag{10 2}$$

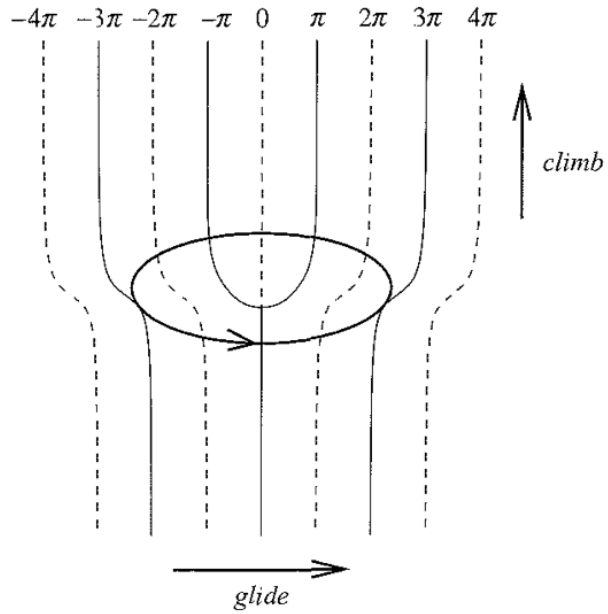


Fig. 10.2 Diagram of a dislocation, showing contours of constant phase. The phase loses 2π in one complete circuit of the closed path around the core. The directions of climb (along roll) and glide (across roll) motions are shown by arrows.

at leading order, where the envelope evolves according to the Newell–Whitehead–Segel equation

$$\frac{\partial A}{\partial T} = \mu A - |A|^2 A + \left(\frac{\partial}{\partial X} - \frac{i}{2}\frac{\partial^2}{\partial Y^2}\right)^2 A \tag{10.3}$$

Figure 10.2 is a diagram of a dislocation. As you can see, the phase of the envelope is undefined at the core of the defect. If we follow a path enclosing the core of the dislocation at a distance, as shown on the diagram, we see that the phase has a discontinuity of 2π, corresponding to the loss of a pair of rolls. In other words, the phase has decreased by 2π by the time the circuit is complete. Since the phase is undefined at the core of the dislocation, the amplitude must be zero there, so that $u(x, y, t)$ is defined everywhere. Dislocations are associated with the discrete translational symmetry, $x \to x + 2\pi$, of the carrier wave e^{ix}: a dislocation locally breaks this symmetry, but interpolates between two solutions that respect it to leading order.

The penta-hepta defect of hexagons (Figure 10.1d) is equivalent to coincident dislocations on two of three component sets of rolls. The pair form a bound state and have opposite jumps in the phase around a path enclosing the core. The

amplitude of both sets of rolls with dislocations is zero at the core, while the amplitude of the third set increases there (see Ciliberto *et al.*, 1990, for further details).

10.2.1 Dislocation motion

Dislocations are not usually stationary: they tend to move through the pattern. Motion along the roll axes is called **climb**: the dislocation moves through states that are related by the continuous translation symmetry along the roll axes. In other words, the relative arrangement of the defect and basic roll pattern is always the same: it is only the position of the defect along the roll axis that varies. Motion across the roll axes, or **glide**, moves through states that are not related by symmetry because the roll structure is periodic in the direction perpendicular to the axes: the translation symmetry in the across-roll direction is discrete, $x \to x + 2\pi$. As a result we expect glide to be resisted by the small scale structure, and the dislocation to be pinned in position in the across-roll direction (Cross & Hohenberg, 1993). A large perturbation or 'kick' in the energy would be needed to get the dislocation moving across the roll axes. However, if there are spatial variations in roll curvature or local wavenumber so that left and right in the across-roll direction are distinguished, we might expect gliding not to be damped because left- or rightward movement might be energetically preferred. If gliding does occur we would not expect the velocity to be constant, but to vary according to the position of the dislocation relative to the roll axes.

10.2.2 Climb

Climbing motion leads to wavelength selection, since as the dislocation climbs upwards the region above the defect shrinks, and the region below, which has a longer wavelength, grows (Figure 10.2). If the dislocation climbs downwards, on the other hand, the region with the shorter wavelength grows.

Energetics can tell us which way the dislocation will climb. The amplitude equation for rolls close to threshold is variational, there being a free energy density associated with rolls at each wavenumber q, as discussed in Chapter 7. In other words, we can write

$$\frac{\partial A}{\partial T} = -\frac{\delta \mathcal{F}}{\delta \bar{A}}, \tag{10.4}$$

where the free energy $\mathcal{F}$ is given by

$$\mathcal{F}(A, \bar{A}) = \int_{\mathcal{D}} f(A, \bar{A}) \mathrm{d}X \, \mathrm{d}Y, \tag{10.5}$$

and where $\mathcal{D}$ is the spatial domain of pattern formation. The free energy density $f(A, \bar{A})$ for the Newell–Whitehead–Segel equation is

$$f(A, \bar{A}) = -\mu|A|^2 + \frac{1}{2}|A|^4 + \left|\left(\frac{\partial}{\partial X} - \frac{i}{2}\frac{\partial^2}{\partial Y^2}\right)A\right|^2, \tag{10.6}$$

and the rate of change of the free energy is given by

$$\frac{\mathrm{d}\mathcal{F}}{\mathrm{d}T} = -2\int_{\mathcal{D}}\left|\frac{\partial A}{\partial T}\right|^2 \mathrm{d}X\,\mathrm{d}Y \leq 0, \tag{10.7}$$

under suitable boundary conditions, as discussed in Section 7.3. This tells us that the free energy must always be decreasing, unless the pattern has stopped evolving, when we have $\mathrm{d}\mathcal{F}/\mathrm{d}T = \partial A/\partial T = 0$. The difference in the free energies of the two sets of rolls will drive the climbing motion of the dislocation, since the pattern with the lower free energy density will be preferred. A roll solution $A = Re^{iqX}$, where $R^2 = \mu - q^2$, has free energy density

$$f = -\frac{1}{2}(\mu - q^2)^2. \tag{10.8}$$

If we have rolls at two different wavenumbers, q_1 and q_2, with corresponding amplitudes, R_1 and R_2, and free energy densities, f_1 and f_2, the difference in the energy densities turns out to be

$$f_1 - f_2 = \left(q_1^2 - q_2^2\right)\left(R_1^2 + R_2^2\right) \tag{10.9}$$

The pattern with lower energy is favoured, and this can now be seen to be the pattern with the lower value of q^2. So the rolls that are closer to critical wavenumber win, and invade the region occupied by the rolls farther from critical. The dislocation climbs towards the rolls with the larger deviation from the critical wavenumber, thereby replacing them with the lower energy pattern.

We can investigate the details of the motion by linearising around a perfect roll pattern at the average wavenumber $q = (q_- + q_+)/2$, where $q \to q_-$ as $Y \to -\infty$ and $q \to q_+$ as $Y \to +\infty$, and where $|q - q_-|, |q - q_+| \ll q$. We shall assume that both q_- and q_+ are positive so that the rolls are stable to the zigzag instability, and that $q_+ > q_-$ so that the dislocation moves upwards. In this calculation we shall largely follow the work of Siggia and Zippelius (1981) and Cross and Hohenberg (1993). We set $A = R_0(1 + R)e^{i(qX+\phi)}$, where $R_0^2 = \mu - q^2$, $0 < R \ll 1$ and $|\nabla\phi| \ll 1$. Now assuming the solution takes the form $R \equiv R(X, Y - vT)$ and $\phi \equiv \phi(X, Y - vT)$, where $v > 0$ is the climbing

velocity of the dislocation, we find

$$-v\frac{\partial R}{\partial Y} = -2R_0^2 R - 2q\frac{\partial \phi}{\partial X} + \frac{\partial^2 R}{\partial X^2} + \frac{\partial^3 \phi}{\partial X \partial Y^2} + q\frac{\partial^2 R}{\partial Y^2} - \frac{1}{4}\frac{\partial^4 R}{\partial Y^4}, \tag{10.10}$$

$$-v\frac{\partial \phi}{\partial Y} = 2q\frac{\partial R}{\partial X} + \frac{\partial^2 \phi}{\partial X^2} - \frac{\partial^3 R}{\partial X \partial Y^2} + q\frac{\partial^2 \phi}{\partial Y^2} - \frac{1}{4}\frac{\partial^4 \phi}{\partial Y^4}, \tag{10.11}$$

which are easily derived from equations (8.6) and (8.7). The standard scalings μ, $\partial/\partial T \sim \epsilon^2$, $\partial/\partial X \sim \epsilon$ and $\partial/\partial Y \sim \epsilon^{\frac{1}{2}}$ have been used, where $0 < \epsilon \ll 1$, so for consistency we must have $v \sim \epsilon^{\frac{3}{2}}$ and therefore the climbing motion of the dislocation is a slow phase-dynamics process.

Following Siggia and Zippelius (1981), we now take the limit $q \to 0$, and for consistency assume that $\partial/\partial X \sim q$, $\partial/\partial Y \sim q^{\frac{1}{2}}$ and $v \sim q^{\frac{3}{2}}$. To leading order we have $R = 0$ and

$$-v\frac{\partial \phi}{\partial Y} = \frac{\partial^2 \phi}{\partial X^2} + q\frac{\partial^2 \phi}{\partial Y^2} - \frac{1}{4}\frac{\partial^4 \phi}{\partial Y^4}. \tag{10.12}$$

We need to add something to this equation to take account of the discontinuity of the phase around the dislocation. If we add

$$-2\pi\frac{\partial(\delta(X - X_d))}{\partial X}\Theta(Y - Y_d), \tag{10.13}$$

where (X_d, Y_d) is the position of the dislocation, δ is the Dirac δ function and Θ is the Heaviside step function, then the phase equation becomes

$$-v\frac{\partial \phi}{\partial Y} = \frac{\partial^2 \phi}{\partial X^2} + q\frac{\partial^2 \phi}{\partial Y^2} - \frac{1}{4}\frac{\partial^4 \phi}{\partial Y^4} - 2\pi\frac{\partial(\delta(X - X_d))}{\partial X}\Theta(Y - Y_d) \tag{10.14}$$

Note that the Heaviside step function $\Theta(Y - Y_d)$ is equal to 1 when $Y > Y_d$ and zero when $Y < Y_d$. Integrating equation (10.14) twice with respect to X and rearranging a bit we find

$$[\phi]_{X-}^{X+} = \int_{X-}^{X+}\left(\int -v\frac{\partial \phi}{\partial Y} - q\frac{\partial^2 \phi}{\partial Y^2} + \frac{1}{4}\frac{\partial^4 \phi}{\partial Y^4}\mathrm{d}X\right)\mathrm{d}X + 2\pi\Theta(Y - Y_d), \tag{10.15}$$

where $[\phi]_{X-}^{X+}$ is the difference in phase between one side of the region containing the dislocation and the other. For $Y > Y_d$ this total phase winding is 2π bigger than for $Y < Y_d$. In other words we have enforced the phase jump of 2π around the dislocation.

The far field structure of a dislocation at rest or moving with constant velocity in the Y direction can be found by solving equation (10.14). Obviously the details of the core will not be captured using this method as we derived this equation by

perturbing around a perfect roll pattern, while at the defect core the pattern is far from perfect and cannot be considered to be a small perturbation to straight rolls.

Far away from the defect core where $Y \gg q^{-\frac{1}{2}}$ and $X \gg q^{-1}$, the fourth derivative in Y is negligible because $\partial/\partial Y \ll q^{\frac{1}{2}}$ and so equation (10.14) reduces to

$$-v\frac{\partial\phi}{\partial Y} = \frac{\partial^2\phi}{\partial X^2} + q\frac{\partial^2\phi}{\partial Y^2} - 2\pi\frac{\partial(\delta(X - X_d))}{\partial X}\Theta(Y - Y_d) \tag{10.16}$$

(Note that we assumed $q > 0$ earlier so that the pattern would be stable to the zigzag instability, so $q^{-\frac{1}{2}}$ is real.) For a stationary dislocation ($v = 0$) at the origin ($X_d = Y_d = 0$) this is solved by

$$\phi = \tan^{-1}\left(\frac{Y}{q^{\frac{1}{2}}X}\right), \tag{10.17}$$

and for a moving defect, assuming that $v \sim q/Y$ as implied by equation (10.16), the solution is given by

$$\frac{\partial\phi}{\partial Y} = \frac{(v\pi)^{\frac{1}{2}}X}{2(qX^2 + Y^2)^{\frac{3}{4}}}e^{-v\{Y+(qX^2+Y^2)^{\frac{1}{2}}\}/2q}. \tag{10.18}$$

This last equation tells us that the dislocation solution will decay exponentially ahead in the direction it is moving ($Y > 0$) and as a power law behind. So two dislocations moving towards each other have a short-range interaction determined by the speed of motion (Siggia and Zippelius, 1981).

Closer to the defect core, we have $Y \ll q^{-\frac{1}{2}}$ and $X \ll q^{-1}$, and now the second derivative in Y is negligible because $q^{\frac{1}{2}} \ll \partial/\partial Y$. In the case of a stationary defect we must now solve

$$0 = \frac{\partial^2\phi}{\partial X^2} - \frac{1}{4}\frac{\partial^4\phi}{\partial Y^4} - 2\pi\frac{\partial(\delta(X - X_d))}{\partial X}\Theta(Y - Y_d), \tag{10.19}$$

and the phase turns out to be

$$\phi = \frac{\pi}{2}\mathrm{sgn}(X)\left[\mathrm{erf}\left(\frac{Y}{\sqrt{2|X|}}\right) + 1\right] \tag{10.20}$$

The particular solutions (10.17), (10.18) and (10.20) can be found using Fourier transforms (see Siggia & Zippelius, 1981).

The analysis is slightly more tractable for anisotropic systems where the phase equation can generally be reduced to the form

$$V\frac{\partial\phi}{\partial\tilde{Y}} + \tilde{\nabla}^2\phi = 0, \tag{10.21}$$

because the spatial derivatives in the original amplitude equation no longer need appear in the special combination $\partial/\partial X - (i/2)\partial^2/\partial Y^2$, and so the X and Y derivatives have the same scaling behaviour. We have dealt with the situation before, when deriving the two-dimensional real Ginzburg–Landau equation (7.52) for rolls in an anisotropic system. Here V is the scaled velocity and $\widetilde{X}$ and $\widetilde{Y}$ are scaled coordinates. Then if $r = (\widetilde{X}^2 + \widetilde{Y}^2)^{\frac{1}{2}}$, the solution is

$$\phi = \tan^{-1}\left(\frac{\widetilde{Y}}{\widetilde{X}}\right), \tag{10.22}$$

for $Vr \ll 1$ and

$$\frac{\partial\phi}{\partial\widetilde{Y}} = \frac{(V\pi)^{\frac{1}{2}}\widetilde{X}}{2r^{\frac{3}{2}}} e^{-V(\widetilde{Y}+r)/2} \tag{10.23}$$

for $Vr \gg 1$. Clearly these are essentially the same solutions as found in the far-field régime of the isotropic case, since there the phase equation takes a similar form. Further details of the analysis for the anisotropic case can be found in Bodenschatz, Pesch and Kramer (1988).

Starting with the solution $A_d(X, Y)$ for a stationary dislocation, the velocity of an isolated defect in an infinite system can be calculated using a method due to Siggia and Zippelius (1981). We follow the Cross and Hohenberg (1993) exposition here. First note that the rate of change of the free energy $\mathcal{F}$ owing to the motion of the dislocation will be given by

$$\frac{d\mathcal{F}}{dT} = v \quad \nabla_d \mathcal{F}_d, \tag{10.24}$$

where v is the velocity of the dislocation, $\nabla_d = (\partial/\partial X_d, \partial/\partial Y_d)$ is the gradient with respect to the position (X_d, Y_d) of the dislocation, and $\mathcal{F}_d = \mathcal{F}(A_d)$. We also have an expression for the rate of change of the free energy in terms of the moving dislocation solution $A_v(X, Y, T)$ from equation (10.7), which is

$$\frac{d\mathcal{F}}{dT} = -2\int_D \left|\frac{\partial A_v}{\partial T}\right|^2 dX\, dY, \tag{10.25}$$

$$= -2\int_D |v \quad \nabla A_v|^2 dX\, dY \tag{10.26}$$

Equating the expressions (10.24) and (10.26) we find that the speed $v = |\boldsymbol{v}|$ of the dislocation is given by

$$v = -\frac{\widehat{v} \cdot \nabla_d \mathcal{F}_d}{2\int_D |\widehat{v} \quad \nabla A_v|^2 dX\, dY}, \tag{10.27}$$

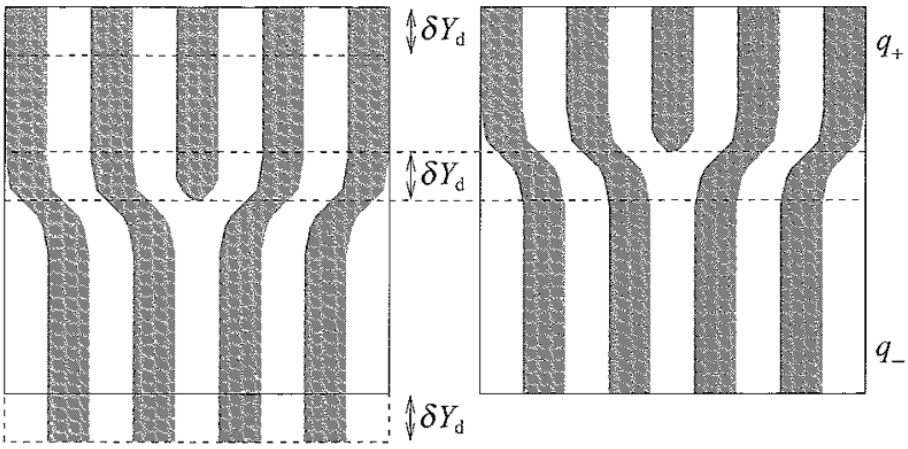


Fig 10 3. The net effect of moving a dislocation along the roll axes by an amount δY_d is to change the wavenumber in a slice of width δY_d from its value at the top of the domain, q_+, to the wavenumber, q_-, at the bottom, or vice versa, depending on the direction of movement. The solid boxes show a fixed small area around the defect core The rolls extend outside the box

where $\hat{v} = v/v$ is a unit vector in the direction of defect motion This is often interpreted as the ratio of the so-called **Peach–Köhler force** to a drag that depends on the amplitude and the direction of motion

In principle it is straightforward to use this method to find the climb velocity of an isolated dislocation in a system close to threshold In this case movement is along the roll axes. Now by symmetry in the Y direction we expect the defect solution to remain the same as it moves, apart from a translation in the Y direction, so to leading order the moving defect solution will take the form $A_v(X, Y, T) = A_d(X, Y - vT)$, where v is the constant climb speed. Equation (10.27) then gives us the value of v:

$$v = -\frac{\frac{\partial \mathcal{F}_d}{\partial Y_d}}{2\int_{\mathcal{D}} \left|\frac{\partial A_d}{\partial Y}\right|^2 \mathrm{d}X\,\mathrm{d}Y} \tag{10.28}$$

To evaluate this expression we need to know the rate of change of the free energy with the position of the defect The net effect of changing the position of the dislocation in the Y direction is to change the wavenumber in a slice of the pattern perpendicular to the roll axes, as can be seen in Figure 10 3. If $q \to q_+$ as $Y \to +\infty$ and $q \to q_-$ as $Y \to -\infty$ then as the dislocation moves from (X_d, Y_d) to $(X_d, Y_d + \delta Y_d)$ the overall free energy of the pattern changes by an amount $\delta \mathcal{F}$ given by

$$\delta \mathcal{F} = (f_{\text{rolls}}(q_-) - f_{\text{rolls}}(q_+))L\delta Y_d, \tag{10.29}$$

where $f_{\text{rolls}}(q)$ is the free energy density for rolls of perturbation wavenumber q (corresponding to $A = R_0 e^{iqX}$), and where L is the length of the pattern-forming region in the X direction. If $(q_+ - q_-)$ is small, we can approximate equation (10.29) by

$$\delta\mathcal{F} = -\frac{\mathrm{d}f_{\text{rolls}}}{\mathrm{d}q}(q_+ - q_-)L\delta Y_{\mathrm{d}}. \tag{10.30}$$

Now $(q_+ - q_-)L$ is the same as $\Delta\phi_+ - \Delta\phi_-$, where $\Delta\phi_+$ is the total phase difference between the right- and lefthand sides of the pattern above the defect, and $\Delta\phi_-$ is the equivalent phase difference below the defect. Looking at Figure 10.2 you can see that this must be equal to the phase-winding around the defect, which is 2π, assuming that $q_+ > q_-$ as in the diagram. So we have

$$\delta\mathcal{F} = -2\pi\frac{\mathrm{d}f_{\text{rolls}}}{\mathrm{d}q}\delta Y_{\mathrm{d}} \tag{10.31}$$

The free energy density for straight rolls, f_{rolls}, is given by substituting the roll solution $A = R_0 e^{iqX}$, where $R_0^2 = \mu - q^2$ into equation (10.6) to find

$$f_{\text{rolls}}(q) = -\frac{1}{2}(\mu - q^2)^2 \tag{10.32}$$

Taking the derivative with respect to q gives

$$\frac{\mathrm{d}f_{\text{rolls}}}{\mathrm{d}q} = 2q(\mu - q^2), \tag{10.33}$$

so we have

$$\frac{\partial F}{\partial Y_{\mathrm{d}}} = -4\pi q(\mu - q^2) \tag{10.34}$$

In the first approximation the obvious thing to do is to use the stationary dislocation solution to work out

$$\int_D \left|\frac{\partial A_{\mathrm{d}}}{\partial Y}\right|^2 \mathrm{d}X\,\mathrm{d}Y \tag{10.35}$$

in order to substitute into equation (10.28) and find the speed of propagation, v. Unfortunately, however, this integral diverges for ϕ given by equations (10.17) and (10.20), and it turns out that we need to solve the full nonlinear phase equation for finite velocity to do the calculation properly. Remember that we started this whole analysis by linearising around a perfect roll solution, and we have not yet added back in any nonlinear terms.

Assuming that the integral is dominated by the contribution from the far field, we have

$$A \approx \sqrt{\mu - q^2}e^{i(qX+\phi)}, \tag{10.36}$$

$$\frac{\partial A_{\mathrm{d}}}{\partial Y} \approx i\sqrt{\mu - q^2}\frac{\partial \phi}{\partial Y}e^{i(qX+\phi)}, \tag{10.37}$$

which together with equations (10.28) and (10.34) gives

$$v \approx \frac{2\pi q}{\int_{\mathcal{D}} \left|\frac{\partial \phi}{\partial Y}\right|^2 \mathrm{d}X\,\mathrm{d}Y}. \tag{10.38}$$

Nondimensionalising equations (10.12) and (10.38), the explicit dependence on q is removed by setting $\widetilde{X} = qX$, $\widetilde{Y} = q^{\frac{1}{2}}Y$ and $\widetilde{v} = vq^{-\frac{3}{2}}$. The nondimensionalised speed, $\widetilde{v}$, is therefore a constant, independent of q, that we determine by evaluating the righthand side of the nondimensionalised version of equation (10.38). This means that v is proportional to $q^{\frac{3}{2}}$. Siggia and Zippelius (1981) used Fourier transforms to approximate the integral using the linear phase equation (10.16) and found a value for the transformed speed, $\widetilde{v}$, nearly twice that deduced from their numerical simulations of the Newell–Whitehead–Segel equation. This is not bad agreement between theory and numerical experiment, given that they completely neglected the nonlinear terms in the phase equation. The theory of defect climb has been refined by inclusion of higher order corrections (Pomeau, Zaleski & Manneville, 1983) and the generation of vertical vorticity by the dislocation, which leads to advection (Tesauro & Cross, 1986). Recently, Walter, Pesch and Bodenschatz (2004) have shown that a balance between the Peach–Köhler force and advection leads naturally to bound dislocation pairs such as are observed experimentally.

The calculations we have carried out so far assume that the system is close to threshold, so that we can use the envelope and carrier wave formalism. In this case we have seen that $v \propto q^{\frac{3}{2}}$ and so we expect a stationary dislocation in the limit $q \to 0$. The existence of the free energy is crucial in the argument we have used. However away from threshold most pattern-forming systems are not variational and there is no free energy. In this case, the velocity can be calculated perturbatively if a stationary dislocation solution is known. There is some experimental and numerical evidence that there is a unique background wavenumber, q_{d}, or a discrete set of wavenumbers, such that an isolated dislocation will be stationary (see, for example, Pocheau & Croquette, 1984, and Tesauro & Cross, 1986). The corresponding defect solution can be used to work out the speed of dislocations at nearby background wavenumbers (Kawasaki, 1984; Tesauro & Cross, 1986). It turns out that the speed depends linearly on $q - q_{\mathrm{d}}$, and there is a cut-off at large

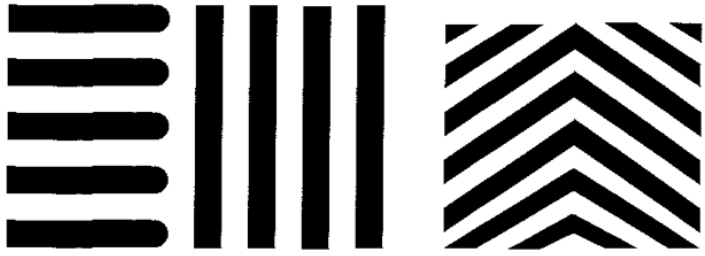


Fig. 10.4 Diagrams of (a) an amplitude grain boundary and (b) a phase grain boundary

distances, owing either to the finite system size or to the velocity itself (Tesauro & Cross, 1986). For further details of the far from threshold case, see Cross and Hohenberg (1993).

10.2.3 Glide

Dislocation glide across roll axes is more complicated. A gliding defect in a background of straight or uniformly curved rolls does not change the overall free energy, and so no movement is expected. For a Swift–Hohenberg model, Kawasaki (1984) shows that in addition to the climbing motion along the roll axes generated by the difference in wavenumbers as $Y \to \pm\infty$, the dislocation velocity also has a component proportional to

$$\hat{z} \times \nabla(\nabla \cdot \boldsymbol{k}) = \hat{z} \times \nabla(\hat{\boldsymbol{k}} \cdot \nabla k + k \nabla \cdot \hat{\boldsymbol{k}}), \tag{10.39}$$

where $\boldsymbol{k}$ is the local wavevector, $k = |\boldsymbol{k}|$ is its modulus and $\hat{\boldsymbol{k}} = \boldsymbol{k}/k$ is a unit vector in the direction of the wavevector. This gives rise to both climb and glide motions. The first term on the righthand side comes from local compression or dilatation of the rolls, and the second from changes in the local roll curvature. So glide is driven by variations in local curvature or second derivatives of the local wavenumber. Glide can also be driven by global anisotropy, for example during convection in an inclined layer (Daniels & Bodenschatz, 2003).

10.3 Amplitude grain boundaries

Grain boundaries occur at the interface between patterns at different orientations. There are two types: **amplitude** and **phase grain boundaries**. A phase grain boundary is a line along which the roll pattern appears to bend sharply, while at an amplitude grain boundary two sets of rolls aligned in different directions meet and overlap for a very short distance; the difference is shown in Figure 10.4, and examples of both types can be seen in Figure 10.1c. In this section we will concentrate

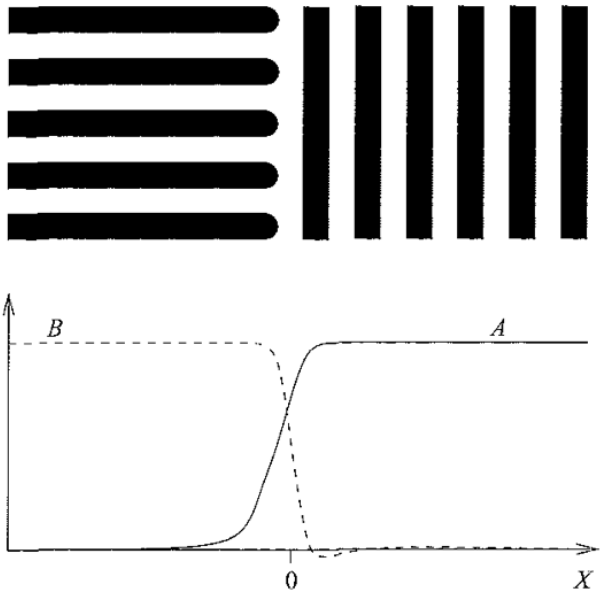


Fig. 10.5. Diagram of an amplitude grain boundary between two roll patterns and the envelopes A and B that describe it. After Manneville (1990).

on amplitude grain boundaries, while phase grain boundaries will be discussed in Chapter 11.

The simplest amplitude grain boundary to analyse is between two roll patterns at right angles, as shown in Figure 10.5. We will investigate this situation following Manneville (1990).

Each set of rolls has its own envelope, so we let A be the envelope corresponding to the rolls with wavevector parallel to the X axis and B be the envelope corresponding to the rolls with wavevector perpendicular to the X axis, and the overall solution is written

$$u(x, y, t) = \epsilon A(X, Y, T)e^{ix} + \epsilon B(X, Y, T)e^{iy} + c.c. + h.o.t., \qquad (10.40)$$

where $0 < \epsilon \ll 1$ is a measure of the distance from the onset of the pattern. From the symmetry of the situation, we assume that there is no Y dependence, so that $A \equiv A(X, T)$ and $B \equiv B(X, T)$. The solution (10.40) is the general form for modulated square patterns, so the amplitude equations are those for bifurcation on a square lattice, with additional spatial derivative terms. Given that there is no Y dependence, the only question is how to introduce the X derivatives. Assuming as usual that the linear growth rate of a Fourier mode $e^{\sigma T + ik\boldsymbol{x}}$, where $\boldsymbol{x} = (x, y)$, is given by

$$\sigma = \mu - \xi_0^2(k^2 - 1)^2 + O([k^2 - 1]^3), \qquad (10.41)$$

where ξ_0 is a real constant and substituting $\boldsymbol{k} = (1 + l, 0)$, where $|l| \ll 1$, to determine the linear part of the equation for A, we find

$$\sigma = \mu - 4\xi_0^2 l^2 + O(l^4) \tag{10.42}$$

On the other hand, substituting $\boldsymbol{k} = (l, 1)$ to look for the behaviour of B gives

$$\sigma = \mu - \xi_0^2 l^4 + O(l^6). \tag{10.43}$$

Inverting the Fourier transform via $2\xi_0 il \to \partial/\partial X$, the factor of $2\xi_0$ introducing a convenient rescaling of X, we deduce that the leading order spatial derivatives are $\partial^2 A/\partial X^2$ and $-(1/4\xi_0^2)\partial^4 B/\partial X^4$. However if, as usual, $X = \epsilon x$, there will be an additional factor of ϵ^2 in front of the term $-(1/4\xi_0^2)\partial^4 B/\partial X^4$. The envelopes therefore evolve according to the equations

$$\frac{\partial A}{\partial T} = \mu A - |A|^2 A - \alpha |B|^2 A + \frac{\partial^2 A}{\partial X^2} + O(\epsilon^2), \tag{10.44}$$

$$\frac{\partial B}{\partial T} = \mu B - |B|^2 B - \alpha |A|^2 B - \frac{\epsilon^2}{4\xi_0^2}\frac{\partial^4 B}{\partial X^4} + O(\epsilon^2), \tag{10.45}$$

where μ and α are real constants, and where the usual scaling $T = \epsilon^2 t$, $X = \epsilon x$, has been adopted, with $\epsilon^2 \mu$ measuring the distance from onset. The small term $-(\epsilon^2/4)\partial^4 B/\partial X^4$ is retained in equation (10.45), while others of the same order are discarded, because it is the leading order spatial derivative term, and can therefore become significant if the amplitude, B, varies in space more rapidly than expected. The scaling with ϵ is included explicitly here, and correspondingly in equation (10.40), to emphasize that the fourth order spatial derivative term appears at a higher order in the expansion. There are also $O(\epsilon^2)$ corrections to the equation for $\partial A/\partial T$, but they are unimportant for our purposes.

The grain boundary could not exist if the rolls were unstable to squares, since then a square pattern would form at the interface and spread throughout the system. Consequently, we must have rolls stable against squares, and hence $\alpha > 1$. We will also assume that the rolls are at critical wavenumber, so that we can take A and B to be real, and that the grain boundary is stationary, so that $\partial/\partial T = 0$.

We shall now follow Manneville (1990) to work out the approximate solutions for the amplitudes A and B. If the position of the grain boundary is $X = 0$, then A is close to zero in $X < 0$ and B is close to zero in $X > 0$, except in a small region very close to the grain boundary. So we can consider there to be two outer regions away from the grain boundary, where one set of rolls dominates the other, and an inner region near the boundary, where the two envelopes are comparable in size. We match the outer solutions at the grain boundary, and check for consistency in the inner region.

In the outer regions, B varies slowly, and so we can neglect the fourth order derivative in equation (10.45). Thus B is slaved to A in space, and either $B = 0$ or

$$B^2 = \mu - \alpha A^2 > 0. \tag{10.46}$$

Equation (10.46) can only hold for $A^2 < \mu/\alpha$. As $X \to \infty$ we expect A to tend to the critical wavenumber roll solution, so $A^2 \to \mu > \mu/\alpha$ (since $\alpha > 1$). Equation (10.46) can therefore only hold in the lefthand ($X < 0$) outer region. Then substituting into equation (10.44) gives

$$0 = -(\alpha - 1)\mu A - (1 - \alpha^2)A^3 + \frac{d^2A}{dX^2}, \tag{10.47}$$

which for the boundary condition $A \to 0$ as $X \to -\infty$ has the solution

$$A_L = \sqrt{\frac{2\mu}{1+\alpha}} \frac{1}{\cosh(\sqrt{\mu(\alpha-1)}(X - X'))}, \tag{10.48}$$

where X' is a real constant.

On the righthand side of Figure 10.5, where A rolls dominate, the only possible solution of equation (10.45) is $B = 0$. Substituting this into equation (10.44) we find

$$0 = \mu A - A^3 + \frac{d^2A}{dX^2}, \tag{10.49}$$

which has the solution

$$A_R = \sqrt{\mu}\tanh(\sqrt{\mu/2}(X - X'')), \tag{10.50}$$

where X'' is a real constant.

Equations (10.47) and (10.49) are second order in space, so we must match the two expressions for both A and dA/dX in the vicinity of the grain boundary. The matching occurs at the point $X = X^*$ at which

$$A_L^2 = A_R^2 = \mu/\alpha, \tag{10.51}$$

because there both solutions of equation (10.45) give $B = 0$. Equations (10.51) comprise two conditions: the first gives the value of $X^* - X'$ and the other fixes $X'' - X'$. The two expressions for dA/dX now match without determining X', which is a free parameter that fixes the position of the grain boundary relative to the chosen coordinate system: in particular we can choose X' such that the grain boundary is at $X = 0$. The outer solution is then given by $A = A_L$, $B^2 = \mu - \alpha A_L^2$ in the left outer region, and $A = A_R$, $B = 0$ in the right outer region.

We must check that this solution does not generate singularities in the internal region around the grain boundary, where B varies rapidly and we can no longer

neglect the term $-(\epsilon^2/4\xi_0^2)\mathrm{d}^4B/\mathrm{d}X^4$ in equation (10.45). Close to $X = X^*$, A and $\mathrm{d}A/\mathrm{d}X$ are continuous, so we can form the Taylor expansion $A = \sqrt{\mu/\alpha} + (X - X_*)c\mu + \cdots$ for some constant c, and substitute into equation (10.45) to find

$$\frac{\epsilon^2}{4\xi_0^2}\frac{\mathrm{d}^4B}{\mathrm{d}X^4} = -2\alpha^{\frac{1}{2}}\mu^{\frac{3}{2}}(X - X_*)cB - B^3. \tag{10.52}$$

Now if we scale $(X - X^*) \sim \epsilon^a$ and $B \sim \epsilon^b$, we find that all the terms are of the same order in ϵ if

$$2 + b - 4a = a + b = 3b, \tag{10.53}$$

which gives $a = 2/5$ and $b = 1/5$. Now the corrections to the amplitude are of order $\epsilon^{\frac{1}{5}}$, and therefore smaller than the outer solution, while lengths in the inner region scale as $\epsilon^{\frac{2}{5}}$, so the corrections are only significant in a narrow strip containing the grain boundary. Consequently there are no inconsistencies in the inner region, and the outer solution is a good approximation to the true one.

Earlier we assumed that the rolls on either side of the grain boundary were at critical wavenumber. In fact, the grain boundary provides a mechanism for wavenumber selection. We can see this by writing $A = Re^{i\phi}$ and $B = Se^{i\psi}$, substituting into equations (10.44) and (10.45) with $\partial/\partial I = 0$ and neglecting the fourth derivative to give

$$0 = R(\mu - R^2 - \alpha S^2) - R\left(\frac{\mathrm{d}\phi}{\mathrm{d}X}\right)^2 + \frac{\mathrm{d}^2R}{\mathrm{d}X^2}, \tag{10.54}$$

$$0 = S(\mu - S^2 - \alpha R^2), \tag{10.55}$$

$$0 = \frac{\mathrm{d}}{\mathrm{d}X}\left(R^2\frac{\mathrm{d}\phi}{\mathrm{d}X}\right) \tag{10.56}$$

According to equation (10.56), the angular momentum $R^2\mathrm{d}\phi/\mathrm{d}X$ is constant in space, but we also have $R \to 0$ as $X \to -\infty$, so we must have $R^2\mathrm{d}\phi/\mathrm{d}X = 0$ everywhere. In other words since R is positive where the A rolls are seen, we must have $\mathrm{d}\phi/\mathrm{d}X = 0$ there and hence ϕ must be constant throughout the region of A rolls. Thus the rolls parallel to the y axis must be at the critical wavenumber. This is very similar to wavenumber selection for rolls on a subcritical ramp, discussed in Chapter 7. It can also be shown that the cross-rolls are at critical wavenumber (see Tesauro & Cross, 1987).

10.4 Domain boundaries between different patterns in systems with a free energy

We will now consider fronts, or domain boundaries, between two different types of pattern that are locally stable, for example rolls and hexagons. When the system is variational, we can use energetics to look at the movement of the front.

First consider the model problem

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\mathrm{d}V}{\mathrm{d}u}, \tag{10.57}$$

where the potential, V, is a function of u only. This instructive example, and a great deal more on fronts besides, can be found in Cross and Hohenberg (1993), whom we shall follow here.

Looking for uniformly translating solutions $u(x,t) \equiv u(\xi)$, where $\xi = x - vt$ with v constant we have

$$\frac{\mathrm{d}^2 u}{\mathrm{d}\xi^2} + v\frac{\mathrm{d}u}{\mathrm{d}\xi} + \frac{\mathrm{d}V}{\mathrm{d}u} = 0 \tag{10.58}$$

Considering ξ to be a time-like variable, this is the equation of motion of a particle of unit mass in a potential $V(u)$, with damping constant v, which in this case can be either positive or negative. Multiplying by $\mathrm{d}u/\mathrm{d}\xi$ and integrating with respect to ξ, we get

$$\frac{1}{2}\left(\frac{\mathrm{d}u}{\mathrm{d}\xi}\right)^2 + v\int\left(\frac{\mathrm{d}u}{\mathrm{d}\xi}\right)^2 \mathrm{d}\xi + V(u) = 0 \tag{10.59}$$

We make the analogy that $(1/2)(\mathrm{d}u/\mathrm{d}\xi)^2$ is the kinetic energy of the particle and $V(u)$ its potential energy. Then writing its total energy E, we find

$$\frac{\mathrm{d}E}{\mathrm{d}\xi} = -v\left(\frac{\mathrm{d}u}{\mathrm{d}\xi}\right)^2 \tag{10.60}$$

So the total energy may increase ($v < 0$), decrease ($v > 0$) or remain constant ($v = 0$).

The stationary solutions of equation (10.58) are $u = u_0$ for constant u_0, and so $\mathrm{d}u/\mathrm{d}\xi|_{u=u_0} = \mathrm{d}^2u/\mathrm{d}\xi^2|_{u=u_0} = 0$. Hence we must have $\mathrm{d}V/\mathrm{d}u|_{u=u_0} = 0$, so the fixed points are the maxima and minima of $V(u)$. Looking at stability, we set $u = u_0 + \hat{u}$, where $|\hat{u}| \ll |u_0|$, and substitute into equation (10.57) to get

$$\frac{\partial \hat{u}}{\partial t} = \frac{\partial^2 \hat{u}}{\partial x^2} + \left.\frac{\mathrm{d}^2 V}{\mathrm{d}u^2}\right|_{u=u_0} \hat{u} + O(\hat{u}^2) \tag{10.61}$$

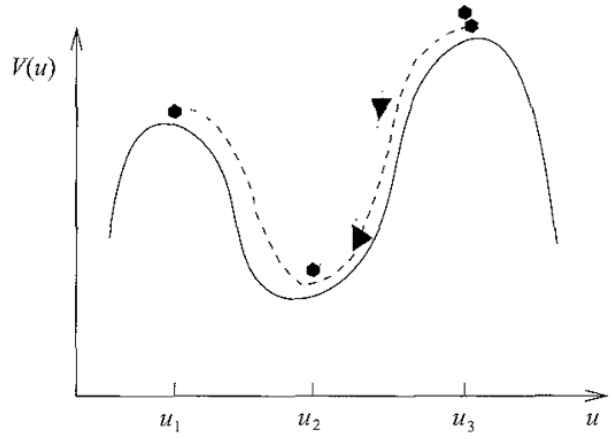


Fig. 10 6 The potential $V(u)$ There are stationary solutions of equation (10 57) at $u = u_1$, u_2 and u_3 The arrows show the direction of increase of the coordinate $\xi = x - vt$, and each path corresponds to a front between two stationary solutions.

The solution $u = u_0$ will therefore be stable if $d^2V/du^2|_{u=u_0} < 0$ and $\partial^2 u/\partial x^2$ is small enough. Consequently, the stable equilibria are the maxima of $V(u)$, while the minima are unstable

The potential shown in Figure 10 6 has two stable maxima and one unstable minimum A front between two stationary solutions is represented by an orbit in phase space that starts at one maximum or minimum and ends at another We can join the righthand maximum at $u = u_3$ to the minimum at $u = u_2$ for a range of values of v. For large $v > 0$, the particle starting from the maximum will be heavily damped, and will roll down to the minimum and stop with no overshoot (dotted path in Figure 10.6) For damping below some threshold value, $v_s > 0$, the particle will overshoot and then return to the minimum Below some critical damping, $v_* > 0$, the particle approaches the minimum through an infinite series of oscillations of decreasing amplitude; the profile, $u(\xi)$, of the leading edge of the front shows a corresponding series of decaying wiggles with increasing ξ For one particular value of the damping, $v = v_{\min} < v_s$, the particle will have just enough energy to reach the other maximum at $u = u_1$ and come to rest there (dashed path in Figure 10.6) The value $v_{\min}$ can be greater or less than v_*; if $v_{\min} > v_*$ the oscillatory fronts will not be seen. Clearly there is a one-parameter family of fronts between the stable state u_3 and the unstable state u_2 with velocities $v_{\min} < v < \infty$, and a unique front between the two stable states u_1 and u_3 with velocity $v_{\min}$.

It would be interesting to know which front is selected out of the whole family of permitted solutions in the stable-to-unstable case For a potential of the form

shown in Figure 10.6, Aronson and Weinberger (1978) proved that, subject to suitable initial conditions, the system selects the front with lowest speed among those that have no overshoot. For a more general evolution equation the matter is very complicated – Cross and Hohenberg (1993) give an overview, and van Saarloos (2003) a comprehensive review. Here we will restrict attention to fronts between locally stable patterns and to energetic results that tell us about the direction of propagation, rather than front speed.

For a front between two constant solutions $u(\xi_i) = u_i, u(\xi_j) = u_j$ with $\xi_i < \xi_j$, we see from equation (10.59) that

$$v = \frac{E(\xi_i) - E(\xi_j)}{\int_{\xi_i}^{\xi_j} \left(\frac{\mathrm{d}u}{\mathrm{d}\xi}\right)^2 d\xi} = \frac{V(u_i) - V(u_j)}{\int_{\xi_i}^{\xi_j} \left(\frac{\mathrm{d}u}{\mathrm{d}\xi}\right)^2 d\xi} \tag{10.62}$$

So we have a rightward moving front $v > 0$ if $V(u_j) < V(u_i)$, and in this case the u_i solution supplants the u_j solution. Conversely we have a leftward moving front $v < 0$ if $V(u_j) > V(u_i)$, and in this case the u_j solution supplants the u_i solution. In other words, the solution corresponding to the greater value of the potential wins. This is not surprising, since we see that

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{-L/2}^{L/2} \left\{ V(u) - \frac{1}{2} \left(\frac{\partial u}{\partial x}\right)^2 \right\} \mathrm{d}x = \int_{-L/2}^{L/2} \left(\frac{\partial u}{\partial t}\right)^2 \mathrm{d}x > 0, \tag{10.63}$$

under suitable boundary conditions, where the domain is $-L/2 \leq x \leq L/2$, and so we expect the system to try to maximise the potential.

In the example shown in Figure 10.6, we see from equation (10.62) that in the special case $V(u_1) = V(u_3)$, we have $v_{\min} = 0$, in other words there is a stationary front between the two stable states. This corresponds to conserving energy in equation (10.58).

10.4.1 Domain boundary between hexagons and rolls

Now we can apply these ideas to the problem of fronts between stable and metastable patterns. We define a metastable state in this context to be a solution that is locally stable to small perturbations, but that has a higher energy than the most stable, lowest energy pattern for a given value of the external stress, μ. If we have a free energy, this will do the job of the potential $V(u)$, and we will easily be able to work out the condition for a stationary domain boundary between two locally stable patterns, by requiring the free energy density to take the same value for both patterns. We require the two patterns to be simultaneously stable to small

perturbations for some value of μ – in other words there must be hysteresis in the system – so we consider hexagons. The hexagonal solution is written in the form

$$u(x, y, t) = A(X, Y, T)e^{ix} + B(X, Y, T)e^{i(-x+\sqrt{3}y)/2} + C(X, Y, T)e^{-i(x+\sqrt{3}y)/2} + c.c. + h.o.t., \tag{10.64}$$

and we shall assume that A, B and C evolve according to

$$\frac{\partial A}{\partial T} = \mu A + a\bar{B}\bar{C} - b|A|^2A - c(|B|^2 + |C|^2)A + \frac{\partial^2 A}{\partial X_1^2}, \tag{10.65}$$

$$\frac{\partial B}{\partial T} = \mu B + a\bar{C}\bar{A} - b|B|^2B - c(|C|^2 + |A|^2)B + \frac{\partial^2 B}{\partial X_2^2}, \tag{10.66}$$

$$\frac{\partial C}{\partial T} = \mu C + a\bar{A}\bar{B} - b|C|^2C - c(|A|^2 + |B|^2)C + \frac{\partial^2 C}{\partial X_3^2}, \tag{10.67}$$

where $X_1 = X$, $X_2 = (-X + \sqrt{3}Y)/2$ and $X_3 = -(X + \sqrt{3}Y)/2$, and where μ, a, b and c are real constants. These are the amplitude equations for modulated hexagons derived in Chapter 7, but with the coefficients g and h set to zero so that there are no spatial derivatives in the quadratic terms. If g and h are nonzero there is no free energy for the system, so we cannot use energetic arguments to determine the direction of movement of fronts, and we will not consider this case here.

In the parameter régime $a > 0$, $c > b > 0$, there is hysteresis in the system: both hexagons at critical wavenumber and the trivial solution are stable in the range $-a^2/4(b + 2c) < \mu < 0$, and both rolls and hexagons at critical wavenumber are stable in the range $a^2b/(b - c)^2 < \mu < a^2(2b + c)/(b - c)^2$, as discussed in Section 5.4. The regions of hysteresis are shown on the bifurcation diagram for patterns on a hexagonal lattice (Figure 5.8).

We will consider the case when all patterns are at critical wavenumber far from the domain boundary, and look for stationary fronts first between up-hexagons, $A = B = C = R_0$, and the trivial solution, and then between up-hexagons and rolls. The free energy functional is given by

$$\mathcal{F}(A, \bar{A}, B, \bar{B}, C, \bar{C}) \equiv \int_D f(A, \bar{A}, B, \bar{B}, C, \bar{C})\mathrm{d}X\,\mathrm{d}Y, \tag{10.68}$$

where $\mathcal{D}$ is the domain of pattern formation and f is the free energy density, defined by

$$\begin{aligned} f(A, \bar{A}, B, \bar{B}, C, \bar{C}) = &-\mu(|A|^2 + |B|^2 + |C|^2) - a(\bar{A}\bar{B}\bar{C} + ABC) \\ &+ (b/2)(|A|^4 + |B|^4 + |C|^4) \\ &+ c(|A|^2|B|^2 + |B|^2|C|^2 + |C|^2|A|^2) \\ &+ \left(\left| \frac{\partial A}{\partial X_1} \right|^2 + \left| \frac{\partial B}{\partial X_2} \right|^2 + \left| \frac{\partial C}{\partial X_3} \right|^2 \right). \end{aligned} \tag{10.69}$$

Under suitable boundary conditions, we have

$$\frac{\mathrm{d}\mathcal{F}}{\mathrm{d}T} = -2 \int_{\mathcal{D}} \left| \frac{\partial A}{\partial T} \right|^2 + \left| \frac{\partial B}{\partial T} \right|^2 + \left| \frac{\partial C}{\partial T} \right|^2 \mathrm{d}X \, \mathrm{d}Y < 0. \tag{10.70}$$

The free energy is always decreasing, so if there is a front between two different stable patterns, then in the absence of boundary effects the front will move so that the pattern with lower free energy density displaces the pattern with higher free energy density. Only if the two patterns have the same energy density will the front remain stationary. For hexagons at the critical wavenumber, the free energy density is

$$f_{\mathrm{h}} = -3\mu R_0^2 - 2aR_0^3 + \frac{3}{2}(b + 2c)R_0^4, \tag{10.71}$$

where R_0 satisfies

$$(b + 2c)R_0^2 - aR_0 - \mu = 0 \tag{10.72}$$

An up-hexagon/trivial solution front would be stationary if $f_{\mathrm{h}} = 0$, which is equivalent to requiring

$$\mu = -\frac{2a^2}{9(b + 2c)}. \tag{10.73}$$

This is in the range $-a^2/4(b + 2c) < \mu < 0$, as required for both hexagons and the trivial solution to be stable.

For rolls at the critical wavenumber the free energy density is

$$f_{\mathrm{r}} = -\mu R_{\mathrm{r}}^2 + \frac{b}{2} R_{\mathrm{r}}^4, \tag{10.74}$$

where $R_r = \sqrt{\mu/b}$ is the roll amplitude. It is easy to show that the free energy densities are equal, $f_h = f_r$, when

$$\mu = \mu_s \equiv \frac{a^2\{b^2 + 3bc + \sqrt{2b}(b+c)^{\frac{3}{2}}\}}{2(b+2c)(b-c)^2}, \tag{10.75}$$

which lies in the range $a^2b/(b-c)^2 < \mu < a^2(2b+c)/(b-c)^2$, where both rolls and hexagons are stable. This is the value of μ at which a roll-hexagon front would be stationary in an infinite layer.

Now consider a moving domain boundary, where to leading order we assume that the speed of propagation, v, is constant. We will consider the symmetric situation in which the direction of propagation is along the x axis, and look for solutions of the form

$$A \equiv A(X - vT), \quad B \equiv B(X - vT), \quad C \equiv C(X - vT). \tag{10.76}$$

The amplitude equations now become

$$-v\frac{dA}{d\xi} = \mu A + a\bar{B}\bar{C} - b|A|^2A - c(|B|^2 + |C|^2)A + \frac{d^2A}{d\xi^2}, \tag{10.77}$$

$$-v\frac{dB}{d\xi} = \mu B + a\bar{C}\bar{A} - b|B|^2B - c(|C|^2 + |A|^2)B + \frac{1}{4}\frac{d^2B}{d\xi^2}, \tag{10.78}$$

$$-v\frac{dC}{d\xi} = \mu C + a\bar{A}\bar{B} - b|C|^2C - c(|A|^2 + |B|^2)C + \frac{1}{4}\frac{d^2C}{d\xi^2}, \tag{10.79}$$

where $\xi = X - vT$, whence we find

$$-2v\left(\left|\frac{dA}{d\xi}\right|^2 + \left|\frac{dB}{d\xi}\right|^2 + \left|\frac{dC}{d\xi}\right|^2\right) = \frac{dE}{d\xi}, \tag{10.80}$$

where

$$E(\xi) = 2\left\{\left|\frac{dA}{d\xi}\right|^2 + \frac{1}{4}\left|\frac{dB}{d\xi}\right|^2 + \frac{1}{4}\left|\frac{dC}{d\xi}\right|^2\right\} - f(A, \bar{A}, B, \bar{B}, C, \bar{C}). \tag{10.81}$$

From this we can see that E is conserved in space across the front for a stationary domain boundary. For patterns at critical wavenumber on either side of the front, the spatial derivatives are zero as $\xi \to \pm\infty$ and so the stationary front must connect two states with the same value of the free energy density, f, as discussed above. The two states on either side of a stationary domain boundary are local minima of the free energy density, and hence are locally stable.

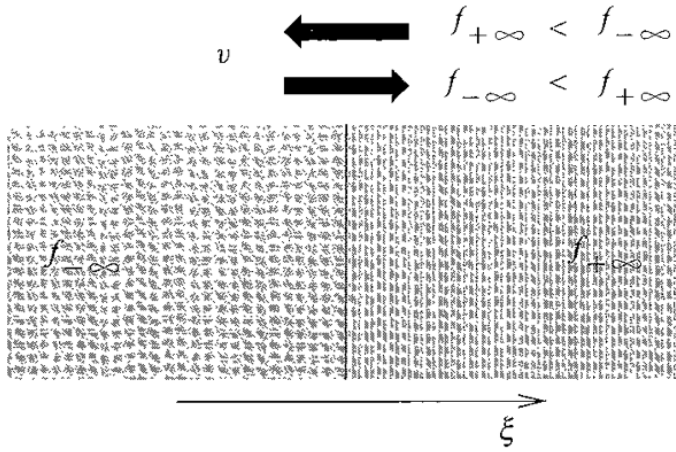


Fig 10.7 The front moves with velocity, v, so that the pattern with lower free energy density, f, invades the other The free energy density as $\xi \to \pm\infty$ is $f_{\pm\infty}$ respectively

Integrating equation (10 80) with respect to ξ gives

$$-2v \int_{-\infty}^{+\infty} \left|\frac{dA}{d\xi}\right|^2 + \left|\frac{dB}{d\xi}\right|^2 + \left|\frac{dC}{d\xi}\right|^2 d\xi = [E]_{-\infty}^{+\infty} \tag{10.82}$$

Again, if the patterns are at critical wavelength far from the domain boundary, then the spatial derivative terms in E vanish, and we have

$$2v \int_{-\infty}^{+\infty} \left|\frac{dA}{d\xi}\right|^2 + \left|\frac{dB}{d\xi}\right|^2 + \left|\frac{dC}{d\xi}\right|^2 d\xi = [f]_{-\infty}^{+\infty} \tag{10.83}$$

This says that if the free energy density is greater as $\xi \to +\infty$ than as $\xi \to -\infty$, the front speed will be positive, and so the pattern at $\xi \to -\infty$ will invade the higher energy pattern at $\xi \to +\infty$, lowering the overall free energy as expected. Conversely, if the free energy density is lower for the pattern at $\xi \to +\infty$, then the front speed v will be negative, and again the lower energy pattern invades the other In other words, the solution corresponding to the lower value of the free energy density always wins (Figure 10.7). This not surprising, since we know that

$$\frac{d\mathcal{F}}{dT} = -2 \int_D \left|\frac{\partial A}{\partial T}\right|^2 + \left|\frac{\partial B}{\partial T}\right|^2 + \left|\frac{\partial C}{\partial T}\right|^2 dX\, dY \leq 0 \tag{10.84}$$

holds, and so the system seeks to minimise the free energy.

In principle we could solve equations (10.77)–(10.79) for A, B and C, and then use equation (10.83) to calculate the speed of propagation, v.

This whole analysis relies on our assumption of the existence of a front moving with constant speed between the two stable patterns. In fact this is hard to prove analytically, but can be verified numerically (see Doelman *et al.*, 2003) for further details and also for proofs of the existence of fronts between stable hexagons and unstable rolls, hexagons or trivial solution in a modified Swift–Hohenberg equation.

The form of the solution (10.76) restricts the patterns to have critical wavenumber in the Y direction, but the theory can be modified to allow off-critical patterns, by setting $A = \widetilde{A}(\xi)e^{iqX}$ and so on, to recover equation (10.80) with the amplitudes replaced by their modified versions $\widetilde{A}$, $\widetilde{B}$ and $\widetilde{C}$, and μ replaced by $\mu - q^2$ wherever it appears. Thus if a front between stable off-critical patterns exists, which is not guaranteed, the two states must have the same free energy density.

We could change the orientation of the domain boundary with respect to the underlying hexagonal carrier wave pattern, which would change the diffusion coefficients in the amplitude equations, and hence the value of v. It has been shown numerically that the dependence of the front velocity on orientation is very small, only about 0.3% (Malomed, Nepomnyashchy & Tribelsky, 1990).

10.5 Energetic considerations for rolls in finite domains

Here we will use the free energy to make some educated guesses about the behaviour of roll patterns close to threshold in finite domains. We want to look at complex geometries and/or large aspect ratio systems where the domain length-scale is much longer than the roll wavelength. Patterns in domains such as these tend to be complicated: the rolls change orientation significantly over the domain and there are lots of defects. Such planforms are often referred to as **textures**: the name comes from the study of liquid crystals, and in our context serves to emphasize the complexity of the patterns.

The free energy is always decreasing, so we expect the pattern to settle down to some local minimum of the free energy. Whether the final state observed corresponds in fact to the global minimum will depend on the size of random fluctuations – thermal noise – in the experimental system. If noise levels are low, the pattern may become 'stuck' in a local minimum or metastable state – in other words a configuration that has higher energy than the global minimum, but that is stable to small perturbations. None the less, we will probably get some idea what kinds of patterns are preferred by working out the contribution to the

free energy from each feature of a general texture, such as grain boundaries, dislocations, curved rolls and so on. According to this approach, due to Cross (1982), the pattern with lowest energy is most likely to be seen. In this section we will follow both Cross (1982) and Manneville (1990) to suggest energetically favourable roll textures.

The orientation of the rolls will vary slowly in space, but we will assume that locally they are described approximately by an envelope and carrier wave formalism, $u(x, y, t) = \epsilon A(X, Y, T)e^{ix} + c.c.$ at leading order, with the envelope evolving according to the Newell–Whitehead–Segel equation,

$$\frac{\partial A}{\partial T} = \mu A - |A|^2 A + \left(\frac{\partial}{\partial X} - \frac{i}{2}\frac{\partial^2}{\partial Y^2}\right)^2 A, \tag{10.85}$$

where $X = \epsilon x$, $Y = \epsilon^{\frac{1}{2}} y$ and $T = \epsilon^2 t$, with $\epsilon^2\mu$ measuring the distance from threshold, and where the x and y directions are defined relative to the local orientation of the rolls. The free energy density for rolls described by this equation is

$$f(A, \bar{A}) = -\mu|A|^2 + \frac{1}{2}|A|^4 + \left|\left(\frac{\partial}{\partial X} - \frac{i}{2}\frac{\partial^2}{\partial Y^2}\right)A\right|^2, \tag{10.86}$$

and in dimensional variables, each term contributes at order ϵ^4. In Chapter 7 we saw that the minimum of the free energy density is $f_{\min} = -\mu^2/2$, corresponding to straight rolls at critical wavenumber. In a finite domain the boundary conditions usually frustrate the development of a uniform pattern of parallel straight rolls, as we shall see, and so defects and bend are introduced, leading to an increase in the overall free energy of the pattern, given by

$$\mathcal{F}(A, \bar{A}) = \int_{\mathcal{D}} f(A, \bar{A}) \mathrm{d}X\, \mathrm{d}Y, \tag{10.87}$$

where $\mathcal{D}$ is the domain.

The roll orientation typically varies significantly over the domain. Strictly speaking, the Newell–Whitehead–Segel equation is not adequate to describe this situation, and we should use a definition of the free energy that captures the large-angled bends of the rolls (see Cross, 1982). In fact we can make reasonable progress within the current framework (Manneville, 1990) and this serves as a simple introduction to the method, which can be adapted straightforwardly to use the more accurate description of the texture.

Contributions to the free energy from the boundary layer close to the walls and from defects will turn out to be important and we will want to distinguish them from contributions made by the bulk of the pattern in the interior of the domain.

This is possible as long as

$$L \gg \epsilon^{-1}, \tag{10 88}$$

where L is a typical lengthscale of the domain, for example the length of a side of a square domain, and ϵ^{-1} is the typical lengthscale of modulations. For convection rolls, the no-slip condition $u(x, y, t) = 0$ at a lateral boundary enforces $A = 0$ there, so there will be a boundary layer over which the amplitude, A, decays to zero from its bulk value. For rolls parallel to the boundary, the relevant spatial derivative in the amplitude equation is $\partial^2 A/\partial X^2$, so the modulation is in X and the boundary layer will have width of order ϵ^{-1}. Each term in the free energy, f, contributes at $O(\epsilon^4)$ and so the increase in the free energy owing to the boundary layer along a length l of wall, will be of order $\epsilon^4 \times \epsilon^{-1} \times l = \epsilon^3 l$. Rolls at an oblique angle to the wall will also contribute an amount of $O(\epsilon^3 l)$ because the dominant modulation will be in X. However, if the rolls are perpendicular to the boundary, then the relevant spatial derivative is $\partial^4 A/\partial Y^4$, the modulation is in Y and so the boundary layer is of width $O(\epsilon^{-\frac{1}{2}})$ and the increase in free energy is $O(\epsilon^{\frac{7}{2}} l)$. Close to threshold we have $0 < \epsilon \ll 1$, so rolls perpendicular to the boundary are much less costly in energetic terms than parallel rolls, and we might expect the perpendicular alignment to dominate. This preference is particular to convection and other systems where boundary conditions enforce a significant modulation of the amplitude close to the walls. The energetic implications of grain boundaries, curvature and defects, that we are about to discuss, are more general.

Consider a grain boundary between parallel and perpendicular rolls. As we saw in Section 10.3, the width of the grain boundary is $O(\epsilon^{-1})$ and so we expect a grain boundary of length l to contribute $O(\epsilon^3 l)$ to the free energy. One possibility for ensuring that rolls are everywhere perpendicular to the boundaries is to introduce a grain boundary in the interior, but since this increases the free energy by the same order of magnitude as would rolls parallel or oblique to the domain boundary, the details of the various alternative configurations will determine which is energetically preferred. Manneville (1990) suggests that since the decaying amplitude does not quite drop off to zero within the inner region of the grain boundary, the overall contribution from a grain boundary will be lower than that from rolls parallel to a boundary, and so a configuration with a grain boundary would be slightly preferable. In practice, the numerical prefactors in front of the $\epsilon^3 l$ would need to be worked out for each individual case. Krishnamurti (1973) observed convection rolls everywhere parallel to the short side of a rectangular box, with cross-roll modulations at each end, suggesting that the grain boundary solution can be selected in experiments with very regular geometry. However, natural patterns

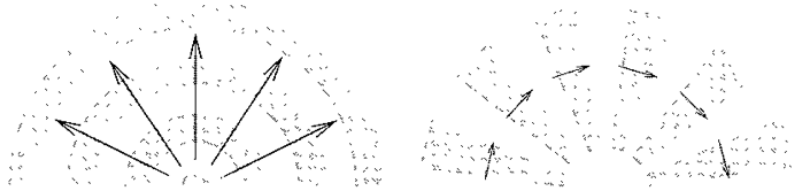


Fig 10 8 Two types of roll curvature: (a) splay and (b) bend The names describe the behaviour of the wavevector, shown in black After Manneville (1990).

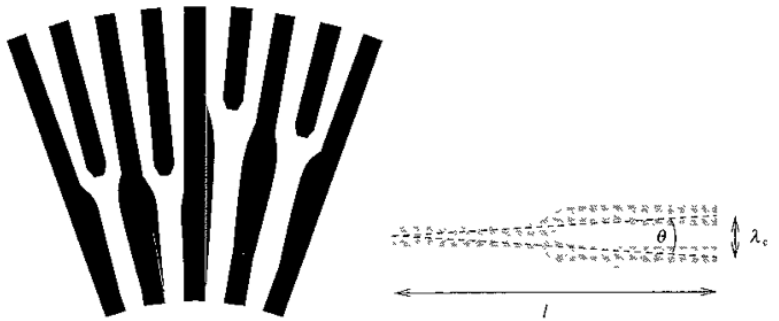


Fig. 10 9. (a) Dislocations nucleate to avoid large wavelength variations in regions of global bend. (b) Each dislocation turns the pattern through an angle, $\theta \sim \lambda_c / l$, where λ_c is the critical wavelength of the rolls, and l is the characteristic lengthscale of the dislocation After Manneville (1990)

usually resolve the problem by introducing large-scale curvature and a range of defects.

Consider the two types of curvature shown in Figure 10.8 In **splay** rolls are curved at constant wavelength, whereas in **bend** the rolls themselves appear almost straight, but there are large variations in wavelength across the bent region Now rolls are only stable with respect to the trivial solution, $u = 0$, if their wavenumber lies within a band of size $O(\epsilon)$ around the critical wavenumber, so the length of the wavevector can vary by no more than $O(\epsilon)$ across the domain. This means that bend, with its large wavelength variations is forbidden Splay, on the other hand, does not change the wavelength to lowest order, and so is a likely configuration.

The system avoids bend by creating dislocations to increase the number of roll pairs in the outer region of the fan (see Figure 10 9a). As bend develops it compresses the inner part of the fan and dilates the rolls in the outer part, in both cases increasing q^2, the square of the deviation in wavenumber, and reducing the amplitude, $|A| = \sqrt{\mu - q^2}$ In order to nucleate a dislocation the amplitude must be zero at the core, so the compression or dilatation leads naturally to the

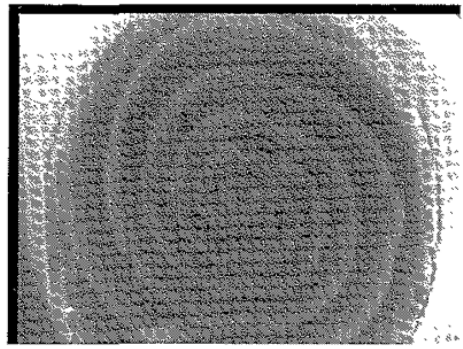


Fig 10.10 Spirals in the Belousov–Zhabotinsky reaction. Image courtesy of and ©Annette Taylor, University of Leeds, August 2004

suppression or creation of a pair of rolls. A single dislocation changes the direction of the pattern by an angle of order λ_c/l, where λ_c is the critical wavelength of rolls and l is the length of the region distorted by the dislocation (see Figure 10.9b). So if $\Delta\theta$ is the total change in angle over the fan, the number of dislocations needed to achieve this is approximately $\Delta\theta/(\lambda_c/l)$. Each dislocation contributes $O(\epsilon^{\frac{5}{2}})$ to the free energy, given by the size of terms in the free energy density, $O(\epsilon^4)$, multiplied by the core area $O(\epsilon^{-1} \times \epsilon^{-\frac{1}{2}})$. The critical wavelength is assumed to be $O(1)$ and if $\Delta\theta$ is also of order unity, then the total free energy contribution from the dislocations will be $O(\epsilon^{\frac{5}{2}}l)$. So a fan of bend and dislocations is less costly than a grain boundary or rolls parallel to a wall, but leads to a higher free energy than either splay, which contributes zero at lowest order, or rolls perpendicular to a boundary.

The free energy analysis suggests then that we should expect a texture with splay curvature allowing rolls to come in perpendicular to the boundaries. Indeed this, with a sprinkling of dislocations, disclinations and grain boundaries, is what is commonly observed (see, for example, Figure 10.17).

10.6 Spirals

Rotating **spiral waves** and their zero-armed counterparts, **targets**, are common in oscillatory and excitable systems, classically in the Belousov–Zhabotinsky chemical reaction (Belousov, 1959; Zaikin & Zhabotinsky, 1970; Figure 10.10) and during the oxidation of carbon monoxide on the surface of a platinum catalyst (see, for example, Nettesheim *et al.*, 1993; Figure 10.12), sometimes described using the FitzHugh–Nagumo model introduced in Chapter 1. They are also seen in the slime mould *Dictyostelium discoideum* when cells aggregate under the influence of the signalling chemical cAMP (Lee, Cox & Goldstein, 1996; Vasiev, Siegert

Fig 10.11 Diagrams of (a) a target or zero-armed spiral, (b) a one-armed spiral and (c) a three-armed spiral

& Weijer, 1997). Spirals and **scroll waves**, the three-dimensional analogues of spirals, also appear in heart muscle during cardiac arrhythmias (see, for example, Winfree, 1998; Panfilov, 1998; Kim *et al.*, 1998). Individually, or in groups, spirals usually fill a large spatial domain, so it is natural to think of them as patterns in their own right, though they can also be regarded as defects in the sense that the phase of the pattern is undefined at the spiral tip. Targets, or zero-armed spirals, are also frequently observed (Figure 10.11a). Spirals and targets in oscillatory or excitable systems emit periodic waves from their core.

In non-oscillatory systems, such as Rayleigh–Bénard convection, spiral and target structures can also be seen; they rotate slowly, but are formed essentially from coiled-up rolls and the evolution is typically slow, controlled by phase diffusion. The waves emanating from the spiral core in this case are annihilated towards the spiral edge by the movement of dislocations on the spiral arms; these structures will be discussed in Section 10.9 on spiral defect chaos.

10.7 Spirals in oscillatory and excitable systems

There are two common types of system in which pulsating spirals and targets are expected to occur: oscillatory, close to an initial Hopf bifurcation from a uniform state, and excitable, where small perturbations to a stable state die away quickly, but larger perturbations persist for a long time, even growing, before the stable state is finally re-established (see the discussion of the FitzHugh–Nagumo model in Chapter 1). In the Belousov–Zhabotinsky reaction, surface catalysis and *Dictyostelium*, spirals and targets are typically a feature of excitable dynamics. However, the complex Ginzburg–Landau equation can support spirals and targets in the oscillatory regime, and it has also been shown that in certain régimes, slime mould patterns are adequately described by the dynamics in the vicinity of a Hopf bifurcation (Hagan & Cohen, 1981). We will start with the complex Ginburg–Landau equation since the analysis here is a bit more tractable.

As shown in Chapter 7, a spatially modulated oscillation in an isotropic, homogeneous system can be written

$$u(x, y, t) = A(X, Y, T)e^{-it} + c.c + h.o.t., \tag{10.89}$$

where the amplitude evolves to leading order according to the complex Ginzburg–Landau equation

$$\frac{\partial A}{\partial T} = \mu A - (1 + i\beta)|A|^2 A + (1 + i\alpha)\nabla^2 A, \tag{10.90}$$

where α and β are real constants. Travelling plane wave solutions are given by

$$A = ae^{iq\,X - i\omega T}, \tag{10.91}$$

$$a^2 = \mu - q^2 > 0, \tag{10.92}$$

$$\omega = \mu\beta + (\alpha - \beta)q^2, \tag{10.93}$$

where $\boldsymbol{X} = (X, Y)$. We define an ***m*-armed** rotating spiral to be a solution of the form

$$A(X, Y, T) = a(r)e^{i\{-\omega T - m\phi + \tilde{\psi}(r)\}}, \tag{10.94}$$

where (r, ϕ) are polar coordinates for $\mathbf{X}$ and the functions $a(r)$ and $\tilde{\psi}(r)$ can in principle be determined by substituting equation (10.94) into the complex Ginzburg–Landau equation and solving the resulting equations. Single-armed and three-armed spirals are shown in Figures 10.11b and c respectively. A target is a zero-armed spiral, so takes the form of equation (10.94) with $m = 0$. Far away from the centre, the spiral or target looks locally like a plane wave, so we have $a^2(r) \to (\mu - q^2)$ and $\partial\tilde{\psi}/\partial r \to q$ as $r \to \infty$. At the tip of the spiral, $r = 0$, the coordinate ϕ is not defined and so neither is the phase of the envelope A. We therefore expect $a(0) \to 0$ as $r \to 0$, so that the envelope remains defined there.

We shall write $A(X, Y, T) = a(X, Y, T)e^{i(-\mu\beta T + \psi(X\,Y\,T))}$, where $a(X, Y, T)$ and $\psi(X, Y, T)$ are real, and look for deviations from the uniformly oscillating state $a = \sqrt{\mu}$, $\psi = 0$ following Kuramoto (1984b). We expect spirals and targets to be driven by phase dynamics, so we suppress any time-dependence in the amplitude, $a(X, Y)$, which is therefore assumed to be slaved to the phase, $\psi(X, Y, T)$. Substituting for A in the complex Ginzburg–Landau equation and separating real and imaginary parts then gives

$$0 = \mu a - a^3 + \nabla^2 a - a|\nabla\psi|^2 - \alpha(2\nabla a \cdot \nabla\psi + a\nabla^2\psi), \tag{10.95}$$

$$a\frac{\partial\psi}{\partial T} = \beta\mu a - \beta a^3 + 2\nabla a \cdot \nabla\psi + a\nabla^2\psi + \alpha(\nabla^2 a - a|\nabla\psi|^2). \tag{10.96}$$

Eliminating $\mu a - a^3$ between these two equations gives the phase equation

$$a\frac{\partial \psi}{\partial T} = (1+\alpha\beta)(2\nabla a \quad \nabla\psi + a\nabla^2\psi) + (\alpha-\beta)(\nabla^2 a - a|\nabla\psi|^2) \qquad (10.97)$$

If the spatial modulation is slow, so that $|\nabla a|/a \ll 1$, then from equation (10.95) we see that

$$a^2 \approx \mu - |\nabla\psi|^2 - \alpha\nabla^2\psi, \qquad (10.98)$$

and so the phase equation (10.96) becomes

$$\frac{\partial \psi}{\partial T} = (1+\alpha\beta)\nabla^2\psi - (\alpha-\beta)|\nabla\psi|^2 \qquad (10.99)$$

to leading order, where a (nonzero) factor of $\sqrt{\mu}$ has been cancelled. This can also be obtained by setting $a \approx \sqrt{\mu}$ in equation (10.97).

We now use equations (10.97) and (10.99) to investigate spiral and target patterns following Kuramoto (1984b) and Cross and Hohenberg (1993).

10.7.1 Externally seeded targets

Targets are often thought to be triggered by the presence of imperfections in the oscillatory medium, so we will first consider these externally seeded target patterns by adding a perturbation to equation (10.99) so that it reads

$$\frac{\partial \psi}{\partial T} = (1+\alpha\beta)\nabla^2\psi - (\alpha-\beta)|\nabla\psi|^2 - g(r), \qquad (10.100)$$

where $g(r)$ is an external inhomogeneity. We must have $(1+\alpha\beta) > 0$ for stability to Benjamin–Feir modes (see Chapter 8), and for now we assume that $(\alpha-\beta)$ is also positive. We have assumed that the seeding perturbation, $g(r)$, is rotationally symmetric and centred at the origin; we shall further assume that it is localised, so that $g(r) = 0$ for $r \geq r_0$, and also positive, with $g(r) > 0$ for $r < r_0$.

Applying the so-called Cole–Hopf transformation $\widetilde{\chi} = e^{-(\alpha-\beta)\psi/(1+\alpha\beta)}$ gives the linear equation

$$\frac{\partial \widetilde{\chi}}{\partial T} = \left\{(1+\alpha\beta)\nabla^2 + \frac{\alpha-\beta}{1+\alpha\beta}g(r)\right\}\widetilde{\chi}, \qquad (10.101)$$

which for solutions of the form $\widetilde{\chi}(r, T) = e^{-\lambda T}\chi(r)$ reduces further to the time-independent Schrödinger equation

$$-(1+\alpha\beta)\nabla^2\chi + V(r)\chi = \lambda\chi, \qquad (10.102)$$

where the potential, $V_{(r)} = -(\alpha-\beta)g(r)/(1+\alpha\beta)$, is both attractive and bounded in space, since $V = 0$ for $r \geq r_0$ and $V < 0$ for $r < r_0$. We need a

solution with $\chi > 0$, since $\widetilde{\chi}$ is positive by definition For potential wells of this sort in two dimensions, there is at least one bound state ($\lambda < 0$) with a positive solution for χ. The value of λ corresponding to this bound state is determined by the shape of the potential well, in other words by the form of $g(r)$

For $r > r_0$ the potential is zero, since $g(r) = 0$ there, and so the Schrödinger equation becomes

$$-(1 + \alpha\beta)\nabla^2\chi = \lambda\chi. \tag{10.103}$$

For $r \gg r_0$ and $\widetilde{q} > 0$,

$$\chi \approx (\widetilde{q}r)^{-1/2}e^{-\widetilde{q}r} \tag{10.104}$$

is an approximate solution of equation (10.103), where $\widetilde{q}^2 = -\lambda/(1 + \alpha\beta)$ So far away from the spiral core we expect

$$e^{-(\alpha-\beta)\psi/(1+\alpha\beta)} \equiv \widetilde{\chi} \approx e^{-\lambda I}(\widetilde{q}r)^{-1/2}e^{-\widetilde{q}r}, \tag{10.105}$$

and taking logarithms

$$-\frac{(\alpha - \beta)\psi}{1 + \alpha\beta} \approx -\frac{1}{2}\ln(\widetilde{q}r) - \lambda I - \widetilde{q}r \approx -\lambda I - \widetilde{q}r \tag{10.106}$$

So the asymptotic form of the phase function ψ as $r \to \infty$ is

$$\psi(X, Y, I) = -\widetilde{\omega}I + qr, \tag{10.107}$$

$$q^2 = -\frac{\lambda(1 + \alpha\beta)}{(\alpha - \beta)^2} > 0, \tag{10.108}$$

$$\widetilde{\omega} = -\frac{\lambda(1 + \alpha\beta)}{\alpha - \beta}, \tag{10.109}$$

which is an approximate plane wave solution with $q > 0$, as required. Adding back in the uniform oscillation $e^{-i\mu\beta T}$ gives the usual dispersion relation (10.93) for travelling plane waves

$$\omega = \mu\beta + \widetilde{\omega} = \mu\beta + (\alpha - \beta)q^2. \tag{10.110}$$

In the absence of perturbations, plane waves are permitted for all values of q The effect of the seeding $g(r)$ is to select a particular wavenumber, $q^2 = -\lambda(1 + \alpha\beta)/(\alpha - \beta)^2$, thus setting the wavelength and frequency of the target pattern. The target oscillates more rapidly than the background, since $\widetilde{\omega}$ is positive, and so the target waves are at higher frequency, as is typical in experiments (Kuramoto, 1984b).

A couple of things about the seeding perturbation are worth mentioning First, if there are several such perturbations in a domain, seeding several targets, there

is no requirement for the form of $g(r)$ to be the same in each case. Typically the perturbations will all be different, and therefore so will the selected wavenumbers and frequencies. Coexistence of targets at different wavenumbers is observed in surface catalysis, where surface imperfections are thought to seed the patterns (Nettesheim *et al*, 1993), though there the dynamics are typically in the excitable regime. In fact it is typical of excitable systems that targets are initiated by spatial inhomogeneities known as **pacemakers**, and so exist for a range of frequencies, while spirals are intrinsic and select a unique period of oscillation (Lee, 1997). Secondly, if $g(r)$ were negative, so that there was a hill in the potential of the Schrödinger equation, rather than a well, there would be no bound states for χ and hence no target patterns.

Target patterns are typically localised with diameters much less than the lengthscale of the domain in which they form. It is likely that many different targets will coexist, and that some of them will meet and influence each other. Each target also interacts with the uniformly oscillating background at its rim, meeting the background state in a shock. We assume the front is at $r = r_s(T) \gg r_0$, with the uniformly oscillating solution $\omega = \beta\mu$, $q = 0$ in the region outside the target, $r > r_s(T)$. If the front were straight rather than curved, the equation for the phase would be

$$\frac{\partial\psi}{\partial T} = (1+\alpha\beta)\frac{\partial^2\psi}{\partial X^2} - (\alpha-\beta)\left(\frac{\partial\psi}{\partial X}\right)^2, \tag{10.111}$$

with boundary conditions $\partial\psi/\partial X \to q$ as $X \to -\infty$ and $\partial\psi/\partial X \to 0$ as $X \to \infty$. The shock solution would then be

$$2(\alpha-\beta)\frac{\partial\psi}{\partial X} = v\left(1+\tanh\left\{-\frac{v}{2(1+\alpha\beta)}(X-vT)\right\}\right) \tag{10.112}$$

with $v = (\alpha-\beta)q$. If the radius of the target pattern is large, then locally the rim will appear straight and this solution will hold approximately, so the rim will move outwards at speed $v_s = (\alpha-\beta)q = (-(1+\alpha\beta)\lambda)^{1/2}$. Within the target itself, but outside the region where the external seeding perturbation is present ($r_0 < r < r_s(T)$), waves propagate outwards with phase velocity

$$v_w = (\epsilon^{-2} + \mu\beta + \tilde{\omega})/q, \tag{10.113}$$

$$= (\epsilon^{-2} + \mu\beta + (\alpha-\beta)q^2)/q, \tag{10.114}$$

where the contribution to the frequency from the carrier wave has been included, and where ϵ^2 measures the distance from onset as usual. The phase velocity is greater than the rim speed, v_s, so wavefronts disappear when they arrive at the rim.

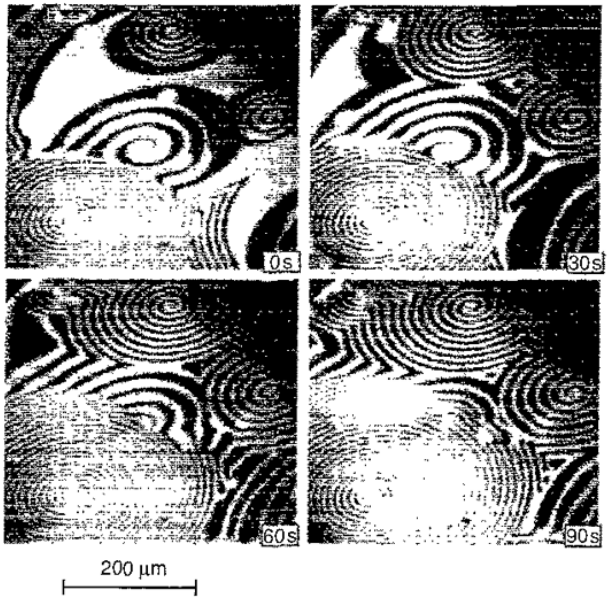


Fig. 10.12. A high frequency and wavenumber spiral consumes another at lower frequency and wavenumber during the oxidation of carbon monoxide on the surface of a platinum catalyst. The time sequence runs top left to bottom right. Reproduced from Nettesheim, S., von Oertzen, A., Rotermund, H.H. and Ertl, G. (1993). Reaction diffusion patterns in the catalytic CO-oxidation on Pt(110): front propagation and spiral waves, *J. Chem. Phys.* **98**, 9977–85, with the permission of the American Institute of Physics.

The outer rims of two target patterns can collide. In general the wavenumbers and frequencies of the two targets will be different as they will be seeded by different random inhomogeneities. There will be a shock between the two targets: the pattern with the higher frequency has the larger wavenumber, q, and hence the higher rim speed, $(\alpha - \beta)q$, so it overruns or consumes the other. This is observed in experiments, as shown in Figure 10.12 reproduced from Nettesheim *et al.* (1993).

Although we assumed that $(\alpha - \beta)$ was positive, we can go through the whole analysis again with $(\alpha - \beta) < 0$. In this case we get targets only for $g(r) < 0$, and now $\tilde{\omega}$ is negative, so that the frequency of oscillation is lower for the targets than for the background state. The wavenumber q, and hence the phase speed, are now also negative, so the phase waves are born at the rim and travel in towards the centre. However the pattern is still a target initiated at the centre by the external inhomogeneity, since the rim and other information-bearing signals still travel outwards at the group velocity $v_s = v_g = (\alpha - \beta)q > 0$.

Hendry *et al.* (2000) also find that targets are stabilised by a localised perturbation, but their approach is different: they allow a radial variation in the external stress, $\mu(r)$, in the complex Ginzburg–Landau equation.

10.7.2 Intrinsic targets

When there are no seeding inhomogeneities, there are no target solutions of the constant amplitude phase equation (10.99). The existence of intrinsic target patterns is controversial, but if we are to find them at all we must look in the phase equation (10.97) that includes spatial variations in the amplitude of the pattern:

$$a\frac{\partial\psi}{\partial T} = (1+\alpha\beta)(2\nabla a \quad \nabla\psi + a\nabla^2\psi) + (\alpha-\beta)(\nabla^2 a - a|\nabla\psi|^2) \qquad (10.115)$$

This allows for the possibility that the centre of the target is a defect, where a varies comparatively rapidly in space. We generalise the transformation we used last time to

$$e^{-\lambda T}\chi(r) = \tilde{\chi}(r,T) = a(r)e^{-(\alpha-\beta)\psi/(1+\alpha\beta)}, \qquad (10.116)$$

and find that χ now satisfies the Schrödinger equation with the modified potential

$$V(r) = (1+\alpha\beta)\left(1+\frac{(\alpha-\beta)^2}{(1+\alpha\beta)^2}\right)\frac{1}{a}\nabla^2 a. \qquad (10.117)$$

Now, the amplitude, a, should be bounded both as $r \to 0$ and as $r \to \infty$. Assuming a power series expansion, plausible asymptotic behaviour might be

$$a(r) \to a_0 + a_k r^k \quad \text{as } r \to 0, \text{ for some } k > 0, \qquad (10.118)$$

$$a(r) \to a_\infty + a_{-1}r^{-1} \quad \text{as } r \to \infty, \qquad (10.119)$$

which corresponds to

$$V(r) \to \kappa_1 \frac{a_k k^2 r^{k-2}}{a_0 + a_k r^k} \quad \text{as } r \to 0, \qquad (10.120)$$

$$V(r) \to \kappa_2 (a_{-1}/a_\infty) r^{-3} \quad \text{as } r \to \infty, \qquad (10.121)$$

where κ_1 and κ_2 are positive constants. Now, if $a_0 = 0$, so that the centre of the target is a defect, as seems likely, then the potential $V(r) \to \kappa_1 k^2 r^{-2}$ as $r \to 0$, so it is strongly repelling at the origin for all k. However as long as a_{-1}/a_∞ is negative, the potential will be attractive ($V < 0$) in the far field. As $r \to \infty$, the amplitude tends to the value $\sqrt{\mu}$, corresponding to the uniform oscillating state, so we have $a_\infty = \sqrt{\mu} > 0$. The coefficient a_1, on the other hand, will be negative if the amplitude approaches its asymptotic value from below, which is expected since $a = \sqrt{\mu - q^2}$ is less than $\sqrt{\mu}$ for a plane wave with wavenumber q. It is

therefore possible, though by no means guaranteed, that spatial modulation of the amplitude could give rise to a potential that has a bound state, and hence a target pattern solution

10.7.3 Spirals

We now turn to spirals. As stated above, an m-armed spiral is a solution of the phase equation (10.97) of the form

$$\psi(X, Y, T) = -\omega T - m\phi + \tilde{\psi}(r), \tag{10.122}$$

$$a(X, Y, T) = a(r). \tag{10.123}$$

We shall assume there is no external seeding. If we make the transformation $\psi(X, Y, T) \to \psi(r, T) - m\phi$ followed by (10.116) there is an additional term in the potential

$$V_m(r) = -\frac{(\alpha - \beta)^2 m^2}{(1 + \alpha\beta) r^2}. \tag{10.124}$$

If the amplitude a is not allowed to vary in space, the total potential is therefore infinitely attracting as $r \to 0$, and so the bound state solution for $\tilde{\psi}$ is highly localised, with $|\lambda| \to \infty$. With $\tilde{\psi}$ changing so rapidly in the vicinity of the origin, the amplitude a also changes over very short distances, which violates the assumption that the amplitude is spatially uniform. Therefore there are no spiral solutions with constant uniform amplitude. If we now introduce spatial dependence of the form

$$a(r) \to \frac{a_m}{r^m} \quad \text{as } r \to 0, \tag{10.125}$$

there is an additional contribution to the potential of the form

$$V_a(r) \to (1 + \alpha\beta)\left(1 + \frac{(\alpha - \beta)^2}{(1 + \alpha\beta)^2}\right)\frac{m^2}{r^2} \quad \text{as } r \to 0, \tag{10.126}$$

that cancels the singularity in V_m leaving a repulsive r^{-2} potential at the origin as in the case of intrinsic targets. If, as before, we assume that the amplitude is bounded as $r \to \infty$, then the leading order contribution to the potential from the space-varying amplitude is $O(1/r^3)$, and so

$$V(r) \to V_m(r) = -\frac{(\alpha - \beta)^2 m^2}{(1 + \alpha\beta) r^2} < 0, \tag{10.127}$$

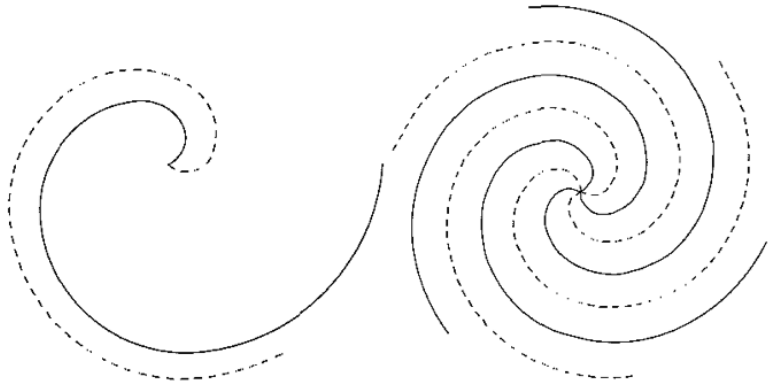


Fig 10 13. Sketches of involute spirals: (a) a one-armed spiral with phase $\psi = r - \phi - T$ and (b) a three-armed spiral with phase $\psi = r - 3\phi - T$. The contour $\psi = 0$ is shown at time $T = 0$ (bold curve) and at a later time $T = T_0 > 0$ (dashed curve).

as $r \to \infty$. Thus the potential is attracting in the far field, and the existence of a bound state, and hence a spiral solution, is plausible, perhaps more so than for an intrinsic target as the far-field attraction is stronger

Since the potential is very small in the far field, the r-dependent part of the phase, ψ, once more takes the approximate form (10.107) there, and the asymptotic form of the whole phase function is

$$\psi(X, Y, T) = -\omega I - m\phi + qr, \tag{10.128}$$

for ω and q constants. Contours of constant phase at a given time, I, take the form of involute spirals (Figure 10.13). The spirals are lefthanded if $q > 0$ and righthanded if $q < 0$. For ω/q positive they rotate clockwise and for ω/q negative the rotation is anticlockwise

Hagan (1982) has shown that one-armed spirals are probably stable and multiarmed spirals probably unstable in this system for $|\beta|$ small enough, while no spirals are stable if $|\beta|$ is large

10.8 Drifting and meandering spirals

Spirals are often observed to **meander** or **drift** (Agladze, Panfilov & Rudenko, 1988; Jahnke, Skaggs & Winfree, 1989; Plesser, Müller & Hess, 1990; Perez-Muñuzuri *et al.*, 1991; Skinner & Swinney, 1991; Nagy-Ungvarai, Ungvarai & Müller, 1993; Nettesheim *et al.*, 1993; Li *et al.*, 1996). The tip of a rigidly rotating spiral traces out a circle, as shown in Figure 10.14a In a frame rotating at the

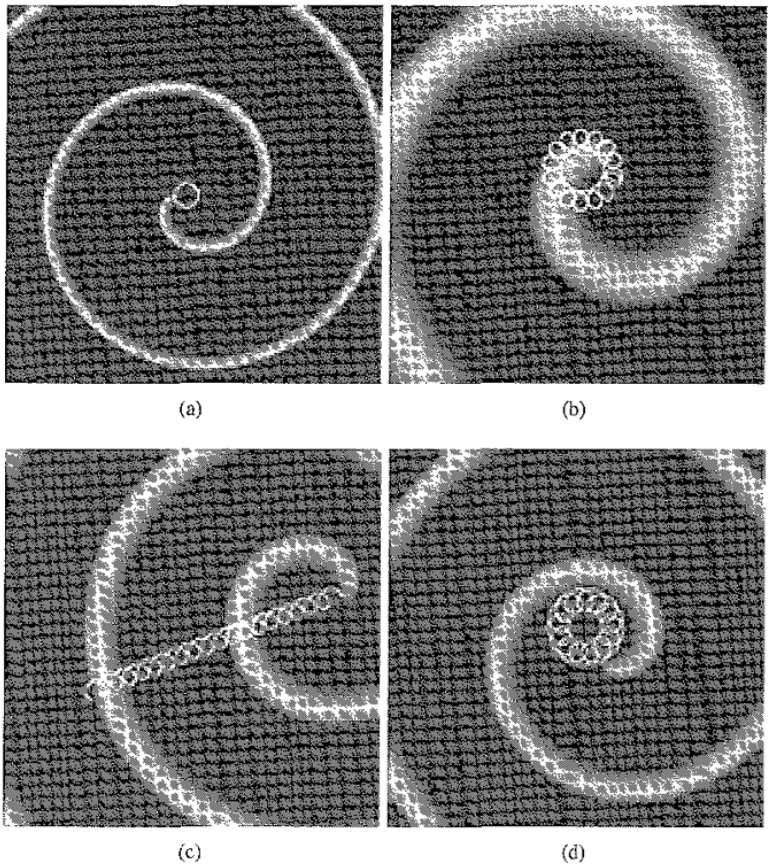


Fig 10 14 Tip paths of spirals: (a) a rigidly rotating spiral with a circular tip path, (b) a meandering spiral with outward petals, (c) a drifting spiral and (d) a meandering spiral with inward petals Pictures courtesy of and ©Björn Sandstede, University of Surrey, July 2004.

same rate as the spiral, the tip position is fixed. In contrast, the tip of a meandering spiral moves in a flower pattern with either inward or outward petals depending on parameter values (Figure 10 14d and b respectively); the tip motion is time-periodic in a corotating frame, and quasiperiodic in the original frame Winfree (1973) noticed drift and writhing of the spiral tip in the Belousov–Zhabotinsky reaction and used the term meander to describe it At the crossover between inward and outward petal meanders, there are spirals that drift off to infinity along a line, about which the tip draws little loops as it goes (Figure 10.14c); in a frame that moves with the drift velocity, the tip motion is once more time-periodic.

Meander and drift are associated with the Euclidean symmetry of the system in which the spirals appear Euclidean symmetry in the plane, $E(2)$, consists of

rotations, reflections and translations. We will only need the rotations and translations to describe spiral motion, so we will use the special Euclidean group, $SE(2)$, that excludes reflections. Both $E(2)$ and $SE(2)$ are noncompact Lie groups, because they include unbounded translations. This causes some technical difficulties, as equivariant bifurcation theory for noncompact groups is not well developed. In Chapter 5 we avoided the problem by restricting solutions to lie on a lattice, reducing the Euclidean symmetry to $H \ltimes T^2$, which is compact. Spirals, however, are not globally periodic in space, so we are stuck with this inconvenient noncompactness. We will look at some simple results, and leave out most of the rigorous justification. If you would like to study the issues in more depth, the article by Sandstede, Scheel and Wulff (1999a) is a good place to start.

We consider a system described by the $SE(2)$-equivariant reaction-diffusion equation

$$\frac{\partial u}{\partial t} = \nabla^2 u + f(u, \mu), \tag{10.129}$$

where $u(\boldsymbol{x}, t) \in \mathbb{R}^n$ is the observable solution, with $\boldsymbol{x} \in \mathbb{R}^2$ and $t \in \mathbb{R}$, where $\mu \in \mathbb{R}^m$ is a vector of bifurcation parameters and where f is typically nonlinear. To identify the types of solution relevant to spiral dynamics, it is helpful to rewrite this as an ordinary differential equation on a function space. In Chapter 5 we did something similar by restricting to a periodic lattice, and projecting the dynamics onto the critical Fourier modes. Here we shall simply assume that it can be done, and that the procedure leads to an $SE(2)$-equivariant mapping, Φ_t, for the time evolution, such that

$$u(t + t_0) = \Phi_t u(t_0) \quad \forall t > 0,\ \forall t_0 \geq 0, \tag{10.130}$$

$$\Phi_t(\gamma u) = \gamma \Phi_t(u), \quad \forall \gamma \in SE(2),\ \forall t > 0, \tag{10.131}$$

where now $u(\boldsymbol{x}, t) \equiv u(t)(\boldsymbol{x})$, so that $u(t)$ is a time-dependent function that acts on the plane. For further details see, for example, Sandstede, Scheel and Wulff (1999b), Wulff (2000) and Wulff (2002).

A rigidly rotating spiral is an example of a **relative equilibrium**, namely a solution, u, such that $u(t) \in \Gamma u(0)$ for all $t \in \mathbb{R}$, for some Lie group, Γ, that may be noncompact. So for a relative equilibrium, evolution in time is the same as evolution along the group orbit, or alternatively, the relative equilibrium is fixed apart from evolution along the group orbit. In a corotating frame, the solution is an equilibrium. Strictly speaking it is the whole group orbit, Γu, that is the relative equilibrium, though it is typical shorthand to refer only to u. In this case $u(t) = \rho_t u(0)$, at a given time, t, for some rotation, $\rho_t \in SO(2) \subset SE(2)$, that depends on t. So time evolution is the same as rotation, as expected.

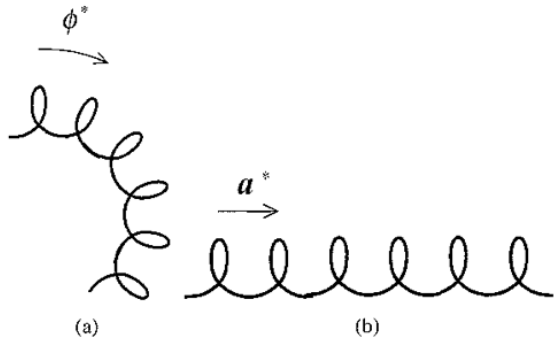


Fig. 10.15 (a) A rotation through an angle ϕ^* maps between consecutive loops in the flower pattern of a meandering spiral tip. (b) For a drifting spiral, the translation a^* is the distance between consecutive loops. After Wulff (2002)

Formally, the relative equilibrium is the group orbit of the rigidly rotating spiral solution under $SE(2)$, and so includes all combinations of translations and rotations of the spiral. All these solutions are essentially the same, in that they behave similarly.

Meandering and drifting spirals are examples of **relative periodic orbits**, which are solutions that satisfy $u(T) \in \Gamma u(0)$ for some time $T > 0$. The minimum T such that this holds is called the **relative period**. This says that after some time, T, the solution returns to a position that is on the group orbit of the initial condition, $u(0)$. The behaviour is periodic because the time-evolution mapping is $SE(2)$-equivariant, and so if $u(T) \equiv \Phi_T u(0) = \gamma^* u(0)$ for some $\gamma^* \in \Gamma \subset SE(2)$ then

$$u(2T) = \Phi_T u(T) = \Phi_T \gamma^* u(0) = \gamma^* \Phi_T u(0) = \gamma^* u(T), \tag{10.132}$$

using the equivariance condition (10.131). So after each relative period, T, the solution has moved along the group orbit by an amount given by γ^*. Strictly speaking, the relative periodic orbit is the time orbit of the group orbit of $u(0)$, namely

$$\mathcal{O} = \{\gamma \Phi_t u(0) : \gamma \in \Gamma, t \in \mathbb{R}\}, \tag{10.133}$$

in other words, all the solutions that can be obtained from $u(0)$ by any combination of time evolution, translation and rotation. This identifies spirals at all points on the flower tip path, and any combination of translations or rotations of these spirals. The solution, $u(0)$, is just a point on the relative periodic orbit, as are the solutions that are obtained from it by time evolution.

In the case of meandering spirals, the group element that provides the period T mapping is a rotation, $\rho^* \in SO(2) \subset SE(2)$, through an angle, ϕ^*, that maps one loop of the flower pattern to the next (Figure 10.15a). For drifting spirals, the relevant group element is the translation, $a^* \in \mathbb{R}^2 \subset SE(2)$, that takes one loop

to the next (Figure 10.15b) In a corotating or comoving frame, the meandering or drifting spiral solutions are periodic orbits, hence the name relative periodic orbit

A rigidly rotating spiral is a rotating wave, a solution where a phase shift, θ, in time is the same as a rotation through $-\theta$ (see Chapter 4) Meandering spirals are sometimes called modulated rotating waves, because there are two frequencies of rotation, one corresponding to the basic rotation of the spiral and the other approximately equal to the frequency of rotation in the corotating frame Similarly, drifting spirals are sometimes called modulated travelling waves, because there is drift superimposed upon the basic spiral rotation

Clearly Euclidean symmetry is key to an understanding of spiral dynamics, so the natural thing to do is to use $SE(2)$-equivariance to write down equations that describe the spiral evolution Barkley (1994) was the first to do this, using a heuristic method Subsequently, the mathematical foundation of his ideas has been put on a more rigorous footing by, for example, Sandstede, Scheel and Wulff (1999a,b) and Wulff (2002), whose articles we shall follow here

First we shall write an element of $SE(2)$ in the form $(\phi, \boldsymbol{a})$, where ϕ is an angle of rotation about $\boldsymbol{x} = \mathbf{0}$, and $\boldsymbol{a} \in \mathbb{R}^2$ is a translation We define the action of $SE(2)$ on solutions, $u(\boldsymbol{x}, t)$, of equation (10 129) by

$$(\phi, a)u(x, t) = u(e^{-i\phi}(x - a), t), \tag{10.134}$$

where the rotation is performed before the translation and where we have identified $\mathbb{R}^2$ with $\mathbb{C}$ so that now $x, a \in \mathbb{C}$ This comes from the standard action of $SE(2)$ on the complex plane and makes the notation particularly neat

A spiral solution that rotates rigidly about the origin, its tip drawing a circle centred on the origin, can be written $u(x, t) = (\phi(t), 0)u_*(x) = u_*(e^{-i\phi(t)}x)$, where $u_*(x)$ is a fixed function describing the shape of the spiral

Generalising briefly to the case where the fixed spiral is rotated by $\phi(t)$ and then translated by $a(t)$, we have a solution

$$u(x, t) = (\phi(t), a(t))u_*(x) = u_*(e^{-i\phi(t)}(x - a(t))). \tag{10.135}$$

The rotation and translation will obey the ordinary differential equations

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = g_1(\phi(t), a(t)), \tag{10.136}$$

$$\frac{\mathrm{d}a}{\mathrm{d}t} = g_2(\phi(t), a(t)), \tag{10.137}$$

where g_1 and g_2 are functions that can in principle be determined from equation (10 129) If $u(x, t)$ is a solution, then so is $(\widetilde{\phi}, \widetilde{a})u(x, t)$ for arbitrary fixed $(\widetilde{\phi}, \widetilde{a}) \in SE(2)$, since the governing equation (10.129) is $SE(2)$-equivariant Therefore

$(\widetilde{\phi}, \widetilde{a})(\phi(t), a(t))u_*(x)$ is a solution and so $(\widetilde{\phi}, \widetilde{a})(\phi(t), a(t))$ must obey the evolution equations (10.136) and (10.137). In Chapter 3 we showed that composition of rotations and translations obeys

$$(\widetilde{\phi}, \widetilde{a})(\phi, a) = (\widetilde{\phi} + \phi, \widetilde{a} + e^{i\widetilde{\phi}}a) \tag{10.138}$$

Substituting the righthand side into equations (10.136) and (10 137) gives

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = g_1(\widetilde{\phi} + \phi(t), \widetilde{a} + e^{i\widetilde{\phi}}a(t)), \tag{10.139}$$

$$e^{i\widetilde{\phi}}\frac{\mathrm{d}a}{\mathrm{d}t} = g_2(\widetilde{\phi} + \phi(t), \widetilde{a} + e^{i\widetilde{\phi}}a(t)), \tag{10 140}$$

and so g_1 and g_2 must satisfy

$$g_1(\phi, a) = g_1(\widetilde{\phi} + \phi, \widetilde{a} + e^{i\widetilde{\phi}}a), \tag{10.141}$$

$$e^{i\widetilde{\phi}}g_2(\phi, a) = g_2(\widetilde{\phi} + \phi, \widetilde{a} + e^{i\widetilde{\phi}}a) \tag{10 142}$$

Now $\widetilde{\phi}$ and $\widetilde{a}$ are arbitrary, and by varying them appropriately the arguments of g_1 and g_2 on the righthand side can take any values, hence the value of g_1 must be independent of its arguments, and so $g_1(\phi, a) = \omega_*$, for some constant ω_* Setting $\widetilde{\phi} = 0$ and varying $\widetilde{a}$, a similar argument shows that $g_2(\phi, a)$ must be independent of a, and so $g_2(\phi, a) \equiv g_2(\phi) = e^{-i\widetilde{\phi}}g_2(\widetilde{\phi} + \phi)$. Thus $g_2(\phi, a) = a_*e^{i\phi}$ for some complex constant a_* So ϕ and a satisfy

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \omega_*, \tag{10.143}$$

$$\frac{\mathrm{d}a}{\mathrm{d}t} = a_*e^{i\phi}, \tag{10 144}$$

and integrating gives

$$\phi(t) = \omega_* t, \tag{10.145}$$

$$a(t) = \frac{ia_*}{\omega_*}(1 - e^{i\omega_* t}), \tag{10 146}$$

using initial conditions $\phi(0) = 0, a(0) = 0$. The equation for $a(t)$ says that the tip of the spiral traces out a circle centred at ia_*/ω_* In coordinates centred at ia_*/ω_* the solution takes the form

$$u(x, t) = (\omega_* t, 0)u_*(x) = e^{-i\omega_* t}u_*(x), \tag{10 147}$$

showing that the spiral rotates rigidly at constant angular velocity ω_* In this coordinate system, we have $a(t) = 0$ for all t, since there is no translation of the tip, only rigid rotation.

It can be shown (Sandstede, Scheel & Wulff, 1997) that as long as $u_*(x)$ is not a stationary solution of equation (10.129) and the eigenvalue spectrum of the operator obtained by linearising the governing equation around u_* satisfies certain conditions, then Σ_*, the isotropy subgroup of u_*, is $\mathbb{Z}_m$ for some $m \in \mathbb{N}$. A spiral with isotropy subgroup $\mathbb{Z}_m$ is m-armed, and has m-fold rotational symmetry

Now we want to look for transitions from a rigidly rotating spiral to meandering or drifting. If we change coordinates to a frame rotating with angular velocity ω_*, the governing equation (10.129) becomes

$$\frac{\partial u}{\partial t} = \omega_* \frac{\partial u}{\partial \phi} + \nabla^2 u + f(u, \mu), \tag{10.148}$$

where ϕ is the angular coordinate in the corotating frame. The rigidly rotating spiral, $u(x, t) = u_*(e^{-i\omega_* t} x)$, is fixed in this frame, so we have

$$\omega_* \frac{\partial u_*}{\partial \phi} + \nabla^2 u_* + f(u_*, \mu) = 0. \tag{10.149}$$

Now writing $u = u_* + \widehat{u}$, where $\widehat{u}$ is small, and linearising gives

$$\frac{\partial \widehat{u}}{\partial t} = \mathcal{L}^* \widehat{u} \equiv \omega_* \frac{\partial \widehat{u}}{\partial \phi} + \nabla^2 \widehat{u} + \left. \frac{\delta f}{\delta u} \right|_{u=u_*} \widehat{u}. \tag{10.150}$$

In order to work out what is happening at the meandering transition, we need to know about the eigenvalue spectrum of $\mathcal{L}^*$. First note that as a result of the Euclidean symmetry $\mathcal{L}^*$ has eigenvalues 0 and $\pm i\omega_*$. The zero eigenvalue is a result of rotational symmetry, since

$$\mathcal{L}^* \frac{\partial u_*}{\partial \phi} = \omega_* \frac{\partial^2 u_*}{\partial \phi^2} + \nabla^2 \frac{\partial u_*}{\partial \phi} + \left. \frac{\delta f}{\delta u} \right|_{u=u_*} \frac{\partial u_*}{\partial \phi}, \tag{10.151}$$

$$= \frac{\partial}{\partial \phi} \left(\omega_* \frac{\partial u_*}{\partial \phi} + \nabla^2 u_* + f(u_*, \mu) \right), \tag{10.152}$$

$$= 0. \tag{10.153}$$

The eigenvector corresponding to the zero eigenvalue is therefore $\partial u_*/\partial \phi$. This is exactly what we would expect, since u_* is a fixed point in the corotating frame, and the symmetry of rotation about the origin carries over from the stationary to the rotating frame, so this is just the zero eigenvalue associated with the continuous rotation symmetry. The translation symmetries, however, are broken in the rotating frame where the spiral solution is an equilibrium, so we don't have zero eigenvalues associated with these. Instead we find a pair of imaginary eigenvalues that come from the transformation into the rotating frame.

Let $x = x_1 + i x_2$, and consider

$$\mathcal{L}^*\left(\frac{\partial}{\partial x_1} \pm i\frac{\partial}{\partial x_2}\right)u_* = \omega_*\frac{\partial}{\partial\phi}\left(\frac{\partial}{\partial x_1} \pm i\frac{\partial}{\partial x_2}\right)u_* + \nabla^2\left(\frac{\partial}{\partial x_1} \pm i\frac{\partial}{\partial x_2}\right)u_* + \left.\frac{\delta f}{\delta u}\right|_{u=u_*}\left(\frac{\partial}{\partial x_1} \pm i\frac{\partial}{\partial x_2}\right)u_*, \tag{10.154}$$

$$= \omega_*\frac{\partial}{\partial\phi}\left(\frac{\partial}{\partial x_1} \pm i\frac{\partial}{\partial x_2}\right)u_* + \left(\frac{\partial}{\partial x_1} \pm i\frac{\partial}{\partial x_2}\right)\left(\nabla^2 u_* + f(u_*, \boldsymbol{\mu})\right), \tag{10.155}$$

$$= \omega_*\frac{\partial}{\partial\phi}\left(\frac{\partial}{\partial x_1} \pm i\frac{\partial}{\partial x_2}\right)u_* - \omega_*\left(\frac{\partial}{\partial x_1} \pm i\frac{\partial}{\partial x_2}\right)\frac{\partial}{\partial\phi}u_*, \tag{10.156}$$

$$= \pm i\omega_*\left(\frac{\partial}{\partial x_1} \pm i\frac{\partial}{\partial x_2}\right)u_*, \tag{10.157}$$

where we have used equation (10.149) to eliminate $\nabla^2 u_* + f(u_*, \boldsymbol{\mu})$ and then $\partial/\partial\phi = x_1\,\partial/\partial x_2 - x_2\,\partial/\partial x_1$. The eigenvectors corresponding to the eigenvalues $\pm i\omega_*$ are therefore $\left(\frac{\partial}{\partial x_1} \pm i\frac{\partial}{\partial x_2}\right)u_*$ respectively.

We will assume that the meandering transition happens at $\boldsymbol{\mu} = \mathbf{0}$ and that close to this point, solutions that are close to a rotation and translation of $u_*(x)$ take the form

$$u(x,t) = (\phi(t), a(t))(u_* + v(t))(x), \tag{10.158}$$

$$= u_*(e^{-i\phi(t)}(x - a(t))) + v(e^{-i\phi(t)}(x - a(t)), t), \tag{10.159}$$

where

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \omega_* + h_1(v, \boldsymbol{\mu}), \tag{10.160}$$

$$\frac{\mathrm{d}a}{\mathrm{d}t} = e^{i\phi}h_2(v, \boldsymbol{\mu}), \tag{10.161}$$

$$\frac{\mathrm{d}v}{\mathrm{d}t} = h_3(v, \boldsymbol{\mu}), \tag{10.162}$$

and where

$$h_1(0, \boldsymbol{\mu}) = h_2(0, \boldsymbol{\mu}) = h_3(0, \boldsymbol{\mu}) = 0. \tag{10.163}$$

The coordinate ϕ is the phase of the spiral wave, while a gives its tip position and v describes changes in its shape. Equation (10.163) says that for the rigidly rotating spiral, $u = u_*$, where $v = 0$, the rotation is at constant velocity and the tip does not translate, as we deduced earlier from equation (10.147). The nonlinear functions h_1, h_2 and h_3 inherit the isotropy of the rigidly rotating spiral according to

$$h_1((\widehat{\phi}, 0)v, \mu) = h_1(v, \mu), \tag{10.164}$$

$$h_2((\widehat{\phi}, 0)v, \mu) = e^{i\widehat{\phi}} h_2(v, \mu), \tag{10.165}$$

$$h_3((\widehat{\phi}, 0)v, \mu) = (\widehat{\phi}, 0)h_3(v, \mu), \tag{10.166}$$

for all rotations $(\widehat{\phi}, 0)$ in the isotropy subgroup of the spiral, $\Sigma_* = \mathbb{Z}_l$. In other words, under a rotation through a multiple of $2\pi/l$, the rate of rotation is unchanged, the direction of translation is rotated, and the shape-change equation is equivariant. This can all be justified rigorously (see, for example, Sandstede, Scheel & Wulff, 1999b).

Barkley (1994) originally proposed that the meandering transition would occur at a Hopf bifurcation from the rigidly rotating spiral. We know that the eigenvalues associated with the ϕ and a variables are zero and $\pm i\omega_*$ respectively, and that the shape-change equation decouples from the others so that the equations have **skew product** structure. We expect that the transition will be driven by a Hopf bifurcation in the shape-change equation, and so we restrict to the case where v is two-dimensional. This gives a five-dimensional centre manifold, just as Barkley had, but the equations are more general.

For a one-armed spiral, the isotropy subgroup consists only of the identity element, so the equivariance condition (10.166) does not restrict the form of h_3 and we can assume that the shape change equation takes Hopf normal form,

$$\frac{\mathrm{d}v}{\mathrm{d}t} = (\mu + i\omega_{\mathrm{H}})v - \alpha|v|^2 v + O(|v|^5), \tag{10.167}$$

where we now have $\mu \in \mathbb{R}$ because we only need one bifurcation parameter for a generic Hopf bifurcation. For μ close to zero there is a small-amplitude periodic solution, $v(t)$, with frequency close to ω_{H}. Substituting this into the equation for $\mathrm{d}\phi/\mathrm{d}t$ we find

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \omega_* + h_1(v(t), \mu) \tag{10.168}$$

Linearising in v and integrating gives

$$\phi(t) = \omega_* t + \widetilde{\phi}(t), \tag{10.169}$$

where $\widetilde{\phi}$ is a small-amplitude, periodic function with frequency close to ω_{H}, and where in the absence of shape changing we have $\phi(0) = 0$. Now substituting both

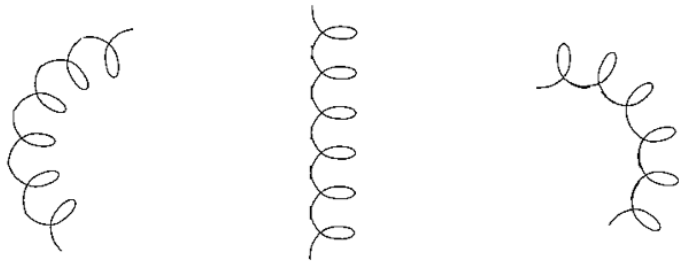


Fig. 10 16. As the resonance $\omega_H = \omega_*$ is approached from either side the radius of the average flower tip path diverges until the average path is a straight line, and the meandering spiral starts to drift instead. The petals on the flower switch from inward at $\omega_H < \omega_*$ to outward at $\omega_H > \omega_*$

$v(t)$ and $\tilde{\phi}(t)$ into the equation for da/dt gives

$$a(t) = a_0 + \int_0^t e^{i(\omega_* \tau + \tilde{\phi}(\tau))} h_2(v(\tau), \mu) d\tau. \tag{10.170}$$

Using the Fourier series expansion

$$e^{i\tilde{\phi}(t)} h_2(v(t), \mu) = \sum_{k \in \mathbb{Z}} b_k e^{ik\omega_H t}, \tag{10.171}$$

we find

$$a(t) = a_0 + \sum_{k \in \mathbb{Z}} b_k \frac{e^{i(\omega_* + k\omega_H)t} - 1}{i(\omega_* + k\omega_H)}, \tag{10.172}$$

as long as there is no resonance such that $\omega_* + k\omega_H = 0$ for any $k \in \mathbb{Z}$. This solution describes a quasiperiodic meandering motion of the spiral tip.

In the resonant case, where $\omega_* + n\omega_H = 0$ for $n \in \mathbb{Z}$ the solution is

$$a(t) = a_0 + b_n t + \sum_{k \in \mathbb{Z}, k \neq n} b_k \frac{e^{i(\omega_* + k\omega_H)t} - 1}{i(\omega_* + k\omega_H)}, \tag{10.173}$$

which describes a spiral drifting off in the direction of b_n, making loops about the line of drift as it goes. Drift is therefore a consequence of resonance between the spiral rotation frequency, ω_*, and the Hopf frequency, ω_H. This was first shown by Barkley (1994) who predicted drift at the resonance $\omega_H = \omega_*$ that marks the switchover from flowers with inward petals ($\omega_H < \omega_*$) to those with outward petals ($\omega_H > \omega_*$). As the resonance is approached, the radius of the average path of the tip diverges so that the flowers open up and approach a drifting pattern (Figure 10.16), as observed by Barkley (1994) in his numerical simulations of a FitzHugh–Nagumo-type model. This is also clear from equation

(10.172) since there is a divergence of the form $(\omega_* + k\omega_{\rm H})^{-1}$ in the tip position as $(\omega_* + k\omega_{\rm H}) \to 0$

Now consider multiarmed spirals with nontrivial isotropy subgroups $\mathbb{Z}_m$ for $m \geq 2$. We know from equation (10.165) that the isotropy is inherited in the equation

$$\frac{{\rm d}a}{{\rm d}t} = e^{i\phi} h_2(v, \mu) \tag{10.174}$$

according to

$$h_2((\widehat{\phi}, 0)v, \mu) = e^{i\widehat{\phi}} h_2(v, \mu), \tag{10.175}$$

for all rotations $(\widehat{\phi}, 0) \in \mathbb{Z}_m$. If there exists a rotation $(\widehat{\phi}, 0) \in \mathbb{Z}_m$, with $\widehat{\phi} \neq 0$, that is also a symmetry of the Hopf eigenfunctions, $v(x)$, in other words a rotation such that

$$(\widehat{\phi}, 0)v(x) \equiv v(e^{-i\widehat{\phi}}x) = v(x), \tag{10.176}$$

then equation (10.175) tells us that

$$h_2(v, \mu) = e^{i\widehat{\phi}} h_2(v, \mu), \tag{10.177}$$

and so $h_2(v, \mu) = 0$ and ${\rm d}a/{\rm d}t = 0$. In this case there can be no drifting or meandering. This says that if the rigidly rotating spiral and the Hopf eigenfunctions have a common symmetry then drifting and meandering do not occur. For example, if the spiral is three-armed it cannot drift if the Hopf eigenfunctions have threefold symmetry.

The theory of meander and drift can be extended to the case of parametric forcing, and to scroll waves, the three-dimensional analogues of spirals (see Sandstede, Scheel & Wulff, 1999a or Wulff, 2002 for an overview).

10.8.1 Superspirals

One can think of a spiral as being created by the emission of waves from its core. Drifting or meandering of the spiral tip should then lead to a Doppler effect, where the wavelength, namely the spacing between consecutive arms, is shortened ahead of the motion and lengthened behind. The overall effect is to introduce a modulation of the spiral wave: this is what the shape-change variable, v, measures. The resultant structures have been named **superspirals**, and have been observed in the Belousov–Zhabotinsky reaction (Perez-Muñuzuri *et al.*, 1991; Li *et al.*, 1996; Zhou & Ouyang, 2000) and during the oxidation of carbon monoxide on a platinum catalyst (Nettesheim *et al.*, 1993). It can be shown (Sandstede & Scheel, 2001) that in the laboratory frame far from the core the meandering and drifting

spirals take the form

$$u = u_\infty(\kappa r + \phi - \omega_* t + \sqrt{\mu}\cos[-\gamma\omega_{\rm gr} t + \gamma r + \phi]), \tag{10.178}$$

$$u = u_\infty(\kappa r + \phi - \omega_* t - \sqrt{\mu}\kappa r\cos[\phi + \alpha]/\omega_{\rm gr}), \tag{10.179}$$

respectively, where $u_*(r,\phi) \to u_\infty(\kappa r + \phi)$ as $r \to \infty$, so the 2π-periodic function u_∞, is the asymptotic shape of the spiral far from the core, and where $\omega_{\rm H} = \omega_* - \gamma\omega_{\rm gr}$, with $\omega_{\rm gr}$ being the group velocity of waves in the far field of meandering spirals, as is clear from equation (10.178), and γ measuring the distance from resonance It can also be shown (Sandstede & Scheel, 2000) that the group velocity, $\omega_{\rm gr}$ is positive, so that information propagates from the core of the spiral outwards, as expected In the solution for drifting spirals, the constant angle α is defined such that the direction of drift is towards $\phi = \pi - \alpha$.

To see the structure of the modulation, we look for level curves of the meandering or drifting spiral Taking meandering first, we set

$$\kappa r + \phi - \omega_* t + \sqrt{\mu}\cos[-\gamma\omega_{\rm gr} t + \gamma r + \phi] = \delta \bmod 2\pi, \tag{10.180}$$

for constant δ. It is the distance between consecutive arms that is modulated, so we examine the quantity $\Delta(r,t)$ that measures the expansion or contraction and satisfies

$$\phi\left(r - \frac{2\pi}{\kappa} + \Delta(r,t)t\right) - \phi(r,t) = 2\pi, \tag{10.181}$$

where $\phi(r,t)$ is given by the solution of equation (10.180), namely

$$\phi(r,t) = \delta - \kappa r + \omega_* t - \sqrt{\mu}\cos[(\omega_* - \gamma\omega_{\rm gr})t + (\gamma - \kappa)r + \delta] + O(\mu), \tag{10.182}$$

for small μ The superspiral structure is most obvious when the compression or contraction is largest, so we shall track the maxima and minima of $\Delta(r,t)$ From equations (10.181) and (10.182) we find

$$\Delta(r,t) = \frac{\sqrt{\mu}}{\kappa}(\phi_1(r - 2\pi/\kappa, t) - \phi_1(r,t)) + O(\mu), \tag{10.183}$$

where

$$\phi_1(r,t) = -\cos[(\omega_* - \gamma\omega_{\rm gr})t + (\gamma - \kappa)r + \delta], \tag{10.184}$$

and so $\partial\Delta/\partial r = 0$ at

$$\frac{\partial\phi_1(r - 2\pi/\kappa, t)}{\partial r} = \frac{\partial\phi_1(r,t)}{\partial r, t} \tag{10.185}$$

For small γ/κ this gives

$$\cos[(\omega_* - \gamma\omega_{\rm gr})t + (\gamma - \kappa)r + \delta] = 0, \tag{10.186}$$

and so

$$\phi + \gamma r - \gamma\omega_{\rm gr} t \approx (\omega_* - \gamma\omega_{\rm gr})t + (\gamma - \kappa)r + \delta = \pm\frac{\pi}{2}. \tag{10.187}$$

This says that the superspiral is itself rotating rigidly with frequency $\gamma\omega_{\rm gr}$. The direction of rotation is correlated with the inward or outward arrangement of petals on the tip path. The meandering spiral solution given in equation (10.178) has motion within each petal with angular velocity ω_*, and motion along the average path, or envelope, with angular velocity $\gamma\omega_{\rm gr}$. Remember that $\omega_{\rm gr}$ is positive, so the direction of rotation along the envelope is given by the sign of γ. If $\omega_* > 0$, so that the basic spiral rotates anticlockwise, then the rotation of the tip within each petal is also anticlockwise. Now if γ is positive ($\omega_{\rm H} < \omega_*$) then the tip also rotates around the envelope anticlockwise, and the petals face inwards. The superspiral rotates anticlockwise too, following its tip. Conversely if γ is negative ($\omega_{\rm H} > \omega_*$) then the tip rotates clockwise around the envelope and the flower has outward petals. In this case the superspiral rotates clockwise. If the basic spiral is rotating clockwise, $\omega_* < 0$, then the inward/outward configurations of the petals is reversed. The crossover between inward- and outward-facing petals agrees with that observed by Barkley (1994), and with experiments (Li *et al*, 1996) that also show the modulational superstructure. Sandstede and Scheel (2001) found agreement between their predictions and direct numerical simulations of the FitzHugh–Nagumo equations.

For drifting spirals, the transformation

$$r \to r + \frac{2\pi n}{|\kappa|}\left(1 + \frac{\sqrt{\mu}}{\omega_{\rm gr}}\cos(\phi + \alpha)\right), \tag{10.188}$$

leads to $\phi \to \phi \mp 2\pi n + O(\mu)$ for $\kappa > 0$ and $\kappa < 0$ respectively, and hence

$$\delta \bmod 2\pi \to (\delta \pm 2\pi n + O(\mu)) \bmod 2\pi = \delta \bmod 2\pi + O(\mu). \tag{10.189}$$

To quadratic order in $\sqrt{\mu}$ the level curves of the spiral are unchanged, so we can identify

$$\Delta(r) = \frac{2\pi n}{|\kappa|}\left(1 + \frac{\sqrt{\mu}}{\omega_{\rm gr}}\cos(\phi + \alpha)\right), \tag{10.190}$$

and so the distance between consecutive arms is compressed by a factor of $(1 - \sqrt{\mu}/\omega_{\rm gr})$ in the direction of drift $\phi = \pi - \alpha$, and expanded by a factor of $(1 + \sqrt{\mu}/\omega_{\rm gr})$ in the opposite direction $\phi = -\alpha$. This constant compression of wavelength ahead and expansion behind is what we would expect of a Doppler effect generated by a source moving at uniform velocity

It is also possible to generate superspirals and supertargets in the absence of drift when $n\omega_* = \omega_H$ for some $n \in \mathbb{Z}$ (Sandstede & Scheel, 2001). These resonances are different from those at which drifting occurs, where $\omega_* = n\omega_H$.

10.8.2 *Spiral break-up and transition to defect-mediated turbulence*

When the Doppler effect caused by meandering spirals is large, neighbouring wavefronts can interfere with each other, leading to break-up of the spiral wave. The **Doppler instability** (Bär *et al.*, 1994; Ouyang, Swinney & Li, 2000) starts near the spiral core where wavefronts break, creating defects – new spiral tips. The process cascades until the whole system is in **defect-mediated turbulence**, a state of spatiotemporal chaos characterised by the creation, annihilation and movement of topological defects (Coullet, Gil & Lega, 1989). It is also possible for spirals to become unstable to a Benjamin–Feir instability (see Chapter 8); the perturbations grow as they propagate outwards from the core, and the spiral breaks up in the far field, forming many defects (Ouyang & Flesselles, 1996). For a mathematical analysis of both core and far-field break-up, see Sandstede and Scheel (2000). Periodic forcing can also destabilise spirals to a turbulent state (Belmonte, Flesselles & Ouyang, 1996).

10.9 Spiral defect chaos

In large-aspect-ratio Rayleigh–Bénard convection, at small Prandtl numbers, straight rolls are observed to compete with a state known as **spiral defect chaos** (Cakmur *et al*., 1997; Melnikov *et al*., 2000). A **large-aspect-ratio domain** is one where the sides are many roll wavelengths long: the terminology comes from convection, where the wavelength is determined by the depth of fluid, and so a system that is extended horizontally for a large number of wavelengths will have a large aspect ratio, in other words the horizontal lengthscale of the system will be much greater than the vertical lengthscale. Highly disordered patterns are more likely in such a system, because most of the domain lies many roll wavelengths away from the walls, so that boundary effects, which tend to impose order (see Section 10.5 above), are weak.

Spiral defect chaos, shown in Figure 10.17, was first observed by Morris *et al* (1993), and is characterised by the complex dynamics of a disordered collection of rotating spirals and targets, travelling dislocations, disclinations, grain boundaries and patches of straight rolls. For an overview of its properties and behaviour, see Bodenschatz, Pesch and Ahlers (2000). The spirals in this case are not oscillatory structures: they usually rotate slowly and are essentially coiled-up convection rolls. Spirals of this type are typical of systems where steady stripes or rolls

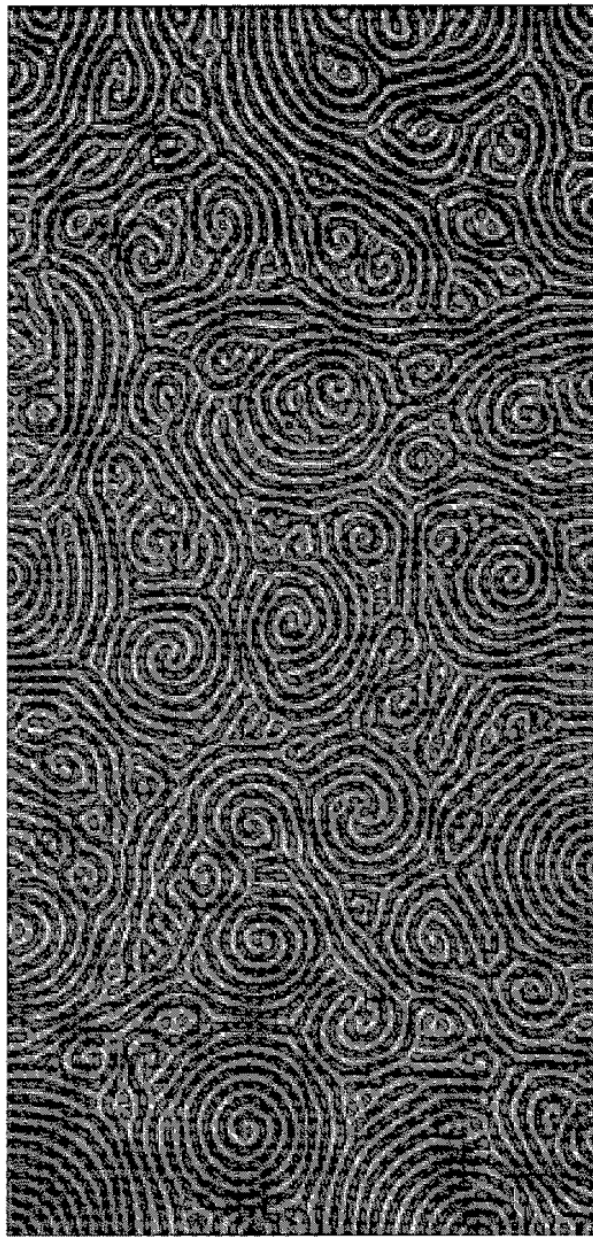


Fig. 10.17. Spiral defect chaos in a Rayleigh–Bénard convection experiment Image courtesy of and ©Nonlinear Phenomena Group, LASSP, Cornell University, August 2004

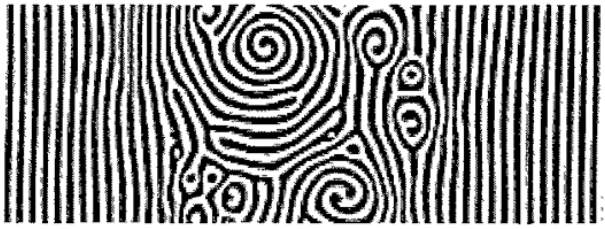


Fig. 10 18 Spiral defect chaos invading a pattern of straight rolls during a numerical simulation of Rayleigh–Bénard convection. Image courtesy of and ©Nonlinear Phenomena Group, LASSP, Cornell University, August 2004

are the expected solution close to threshold. Giant steady spirals can also occur in Rayleigh–Bénard convection. The waves that propagate from the core as a result of their rotation are removed by dislocations travelling along the spiral arms (Plapp *et al.*, 1998); this is a slow phase-diffusion process. Targets can also be seen in these systems, and in this context they are often called foci. Again they are essentially steady rather than oscillatory, though roll rings can be added or removed though the nucleation and movement of dislocations, and targets can be destabilised in the **focus instability** (Newell, Passot & Souli, 1990).

Close to the onset of convection, stable steady straight rolls are predicted theoretically. In fact, for large-aspect-ratio, small-Prandtl-number convection, straight rolls are only seen if the initial conditions are specially prepared to favour them (see, for example, Cakmur *et al.*, 1997); for general initial conditions, a weakly chaotic time-dependent state emerges close to onset, while above a threshold value, r_s, of the reduced Rayleigh number, spiral defect chaos is seen. The threshold, r_s, decreases with increasing aspect ratio and with decreasing Prandtl number. Above the threshold both rolls and spiral defect chaos are stable over a wide range in r, though straight rolls can only be achieved starting from special initial conditions as before, while random initial conditions lead to spiral defect chaos. Toward the lower r end of the bistable regime, patches of nearly straight rolls grow and compete with regions of spiral defect chaos, while for higher values of the reduced Rayleigh number, spiral defect chaos typically invades a pre-prepared straight roll state via an almost straight front moving at approximately constant speed (Cakmur *et al.*, 1997; Melnikov *et al.*, 2000). Figure 10 18 shows spiral defect chaos invading rolls from the centre of the domain during a numerical simulation of Rayleigh–Bénard convection.

The spiral defect chaos state is coupled to a mean drift generated by the roll curvature. Numerical simulations (Chiam *et al.*, 2003) suggest that the mean flow is essential for the development of spiral defect chaos, and that if that flow is quenched the pattern collapses into patches of stationary rolls with sharp angular bends and joins. The average wavenumber of spiral defect chaos lies within

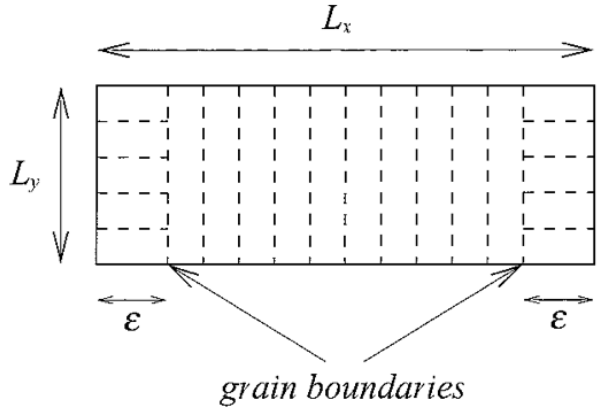


Fig 10.E1 Diagram of Rayleigh–Bénard convection cell for Exercise 10.3

the stable Busse balloon for straight rolls, but there is a tail of high wavenumbers that lies above the skew-varicose boundary, and defects are observed to form through the skew-varicose instability described in Chapter 8. An even smaller tail of wavenumbers is found to lie below the cross-roll instability boundary (see Chapter 8), but cross-roll events are rarely observed (Egolf, Melnikov & Bodenschatz, 1998).

Exercises

10.1 Find the direction of movement of a front between locally stable off-critical up-hexagons and locally stable off-critical rolls if both patterns have the same perturbation wavenumber, q, and the system is described by equations (10.65)–(10.67) in the parameter régime $a > 0$, $c > b > 0$, and $\mu = \mu_s$, where μ_s defined in equation (10.75) is the value at which a front between critical up-hexagons and critical rolls is stationary. Does your answer depend on the value of q? [Hint: work out which pattern has lower free energy on either side of $\mu = \mu_s$ by considering a value of μ at which one of the free energies takes a particularly simple form.]

10.2 Consider rolls, $u(x, y, t) = A(X, Y, T)e^{ix} + c.c. + h.o.t.$, where the amplitude evolves according to

$$\frac{\partial A}{\partial T} = \mu A - |A|^2 A + a\left(\frac{\partial A}{\partial X} - \frac{i}{2}\frac{\partial^2 A}{\partial Y^2}\right)^2 \bar{A}$$

$$+(1 + a|A|^2)\left(\frac{\partial}{\partial X} - \frac{i}{2}\frac{\partial^2}{\partial Y^2}\right)^2 A.$$

If there is an isolated dislocation at the origin and if $A \to R_1 e^{iq_1 X}$ as $Y \to -\infty$ and $A \to R_2 e^{iq_2 X}$ as $Y \to +\infty$, where $q_1^2 \neq q_2^2$, find a relationship between q_1 and q_2 such that the dislocation is expected to be stationary on energetic grounds.

10.3 Consider the rectangular Rayleigh–Bénard convection cell shown in Figure 10.E1. The sides are of length L_x and L_y, where L_y is the shorter side. The cell is filled with rolls at critical wavenumber that are everywhere parallel to the short sides of the box, except within a distance ϵ of the short ends, where they come in perpendicular to the short sides. There are grain boundaries at a distance ϵ from either short end separating the regions of rolls parallel and perpendicular to the short sides. Assume that a grain boundary contributes 90% as much per unit length to the free energy as rolls parallel to a wall. If the configuration in Figure 10.E1 is to be energetically favoured, work out:

(a) an approximate upper bound for ϵ/L_y in terms of the energetic costs per unit length of rolls perpendicular and parallel to a wall, and

(b) an approximate upper bound on the cost per unit length of rolls perpendicular to a wall as a fraction of the cost of rolls parallel to a wall.

10.4 Work out the form of the tip path, $a(t)$, for the meandering one-armed spiral described by equations (10.160), (10.161) and (10.167) with $h_1(v, \mu) = 0$ and $h_2(v, \mu) = \gamma_0 + v$, where γ_0 is a complex constant.

11

Large-aspect-ratio systems and the Cross–Newell equation

In Chapter 7 we allowed patterns to deviate slightly from a regular lattice by permitting modulations on long scales. However, in a large-aspect-ratio system that can accommodate a large number of pattern wavelengths in all directions, the size and orientation of the pattern will typically change slowly in space and time. In spiral defect chaos, for example, you tend to see patches of rolls that look quite regular locally, but in fact are curved with a large radius of curvature (Figure 10.17). Fingerprints, though stationary, also look like stripes that vary slowly in orientation over a large domain. The ridges in fingerprints are believed to form through the buckling of the lower layer of the skin; recently Kuecken (2004) has derived roll-hexagon amplitude equations from a buckling model of fingerprint formation, suggesting that the analysis of fingerprints as a pattern-forming system may be valid.

Obviously we can't describe patterns in a large-aspect-ratio system by assuming that they lie almost on a lattice, since they clearly don't. However, far from onset in the fully nonlinear regime, we can use the slowness with which the patterns evolve in time and space to develop an asymptotic description of them. The full nonlinearity is a requirement of the theory, so we will lose the small parameter measuring the distance from onset that we used previously to derive amplitude equations, but the slow rates of change will give us a new small parameter to work with.

The theory presented in this chapter was originally developed by Cross and Newell (1984) and later expanded by Passot and Newell (1994). Here we follow their treatment of the problem quite closely. Defects as solutions of the Cross–Newell equations were studied extensively by Newell *et al.* (1996), Ercolani *et al.* (2000) and Ercolani *et al.* (2003); their papers consider the issue in much more detail than this chapter will attempt, and are a natural starting point for those who wish to study the matter further.

11.1 Fully nonlinear patterns in large-aspect-ratio boxes

We will consider roll or stripe patterns as this is the simplest case. The plan is to describe the rolls by their local wavevector. Since the pattern is fully nonlinear the roll amplitude will be slaved to the wavevector, and so once we know how the wavevector wiggles around in space and time, we will have specified the pattern completely.

We assume that fully developed rolls in a homogeneous isotropic system can be described by a stationary solution $u(x, y, t) = u_0(x, y)$ of an equation

$$\frac{\partial}{\partial t}u(x, y, t) = \mathcal{L}u(x, y, t) + \mathcal{N}(u(x, y, t)), \tag{11.1}$$

where $\mathcal{L}$ is a linear differential operator and $\mathcal{N}$ is a nonlinear operator.

As stated above, the rolls are described by their local wavevector, $\boldsymbol{k}$, which is associated with a local phase, θ, according to $\boldsymbol{k} = \nabla\theta$ (So for x rolls with $|\boldsymbol{k}| = 1$ the phase is $\theta = x$.) For fully nonlinear rolls, the amplitude is determined adiabatically from the phase θ except in the vicinity of certain defects, where the amplitude is small, or when the driving stress parameter of the system is close to the critical value for pattern formation so that the amplitude is small everywhere.

Since we are analysing a large-aspect-ratio system, we make the assumption that the wavevector of the pattern will typically change slowly in space and time. To describe these changes, we introduce a large-scale phase $\Theta = \epsilon\theta$, where $0 < \epsilon \ll 1$ is a small parameter that measures the slowness of the variations. In a convection experiment, ϵ would be the inverse aspect ratio of the box; in general ϵ is the ratio of the basic pattern wavelength to the system size. Close to a defect the pattern orientation varies rapidly, so in the presence of defects ϵ is defined to be the inverse of the average distance between them. We also define slow space- and timescales $X = \epsilon x$, $Y = \epsilon y$, and $T = \epsilon^2 t$. The x and y scalings are identical since we want to study patterns that have no preferred direction over the domain as a whole, and so we must respect rotational symmetry. The local wavevector is then given by $\boldsymbol{k} = \nabla_x\theta = \nabla_X\Theta$, where

$$\nabla_x \equiv \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right), \quad \nabla_X \equiv \left(\frac{\partial}{\partial X}, \frac{\partial}{\partial Y}\right) \tag{11.2}$$

The roll solution is now considered to be a function of the phase θ and the slow space- and timescales, such that $u \equiv u(\theta; X, Y, T)$. Hence the space and time

derivatives of u are given by

$$\nabla_x u(\theta; X, Y, T) = (\nabla_x \theta)\frac{\partial u}{\partial \theta} + (\nabla_X u \cdot \nabla_x)(X, Y)$$
$$= \left(\boldsymbol{k}\frac{\partial}{\partial \theta} + \epsilon \nabla_X\right) u(\theta; X, Y, T), \tag{11.3}$$
$$\frac{\partial}{\partial t} u(\theta; X, Y, T) = \frac{\partial \theta}{\partial t}\frac{\partial u}{\partial \theta} + \frac{\mathrm{d}T}{\mathrm{d}t}\frac{\partial u}{\partial T}$$
$$= \left(\epsilon \frac{\partial \Theta}{\partial T}\frac{\partial}{\partial \theta} + \epsilon^2 \frac{\partial}{\partial T}\right) u(\theta; X, Y, T) \tag{11.4}$$

We now expand the solution u and the linear operator $\mathcal{L}$ in powers of ϵ so that

$$u = u_0 + \epsilon u_1 + \epsilon^2 u_2, \tag{11.5}$$
$$\mathcal{L} = \mathcal{L}_0 + \epsilon \mathcal{L}_1 + \epsilon^2 \mathcal{L}_2. \tag{11.6}$$

Then at leading order in ϵ we have

$$0 = \mathcal{L}_0 u_0 + \mathcal{N}(u_0) \tag{11.7}$$

Neither $\mathcal{L}_0$ nor $\mathcal{N}$ depend on time, nor on the long scales X and Y, so we consider u_0 to depend only on θ and the local wavenumber $k = |\boldsymbol{k}|$. Thus equation (11.7) slaves the amplitude, $|u_0|$, to k, via $\boldsymbol{k} = \nabla_x \theta$. Owing to translation symmetry, $u_0(\theta + \delta\theta)$ must also be a solution of equation (11.7) for constant $\delta\theta$ and so we must have

$$0 = \mathcal{L}_0 u_0(\theta + \delta\theta) + \mathcal{N}(u_0(\theta + \delta\theta)), \tag{11.8}$$

and hence

$$0 = \left(\mathcal{L}_0 + \frac{\delta \mathcal{N}}{\delta u}\bigg|_{u=u_0}\right)\frac{\partial u_0}{\partial \theta}. \tag{11.9}$$

At $O(\epsilon)$ we find

$$\frac{\partial \Theta}{\partial T}\frac{\partial u_0}{\partial \theta} = \mathcal{L}_0 u_1 + \mathcal{L}_1 u_0 + \frac{\delta \mathcal{N}}{\delta u}\bigg|_{u=u_0} u_1. \tag{11.10}$$

Now we know that the operator $\mathcal{L}_0 + \delta\mathcal{N}/\delta u|_{u=u_0}$ is singular, since equation (11.9) tells us that $\partial u_0/\partial\theta$ is in its null space, that is the operator acts on $\partial u_0/\partial\theta$ to give zero. Then according to the Fredholm alternative there is a solvability condition on equation (11.10). We make the assumptions that u_0 is chosen to be 2π periodic in θ, that u_1 is also required to be so, and that $\mathcal{L}_0$ is self-adjoint. Under

these circumstances we have

$$\int_0^{2\pi} \frac{\partial u_0}{\partial \theta} \left(\mathcal{L}_0 + \left. \frac{\delta \mathcal{N}}{\delta u} \right|_{u=u_0} \right) u_1 \, \mathrm{d}\theta = \int_0^{2\pi} u_1 \left(\mathcal{L}_0 + \left. \frac{\delta \mathcal{N}}{\delta u} \right|_{u=u_0} \right) \frac{\partial u_0}{\partial \theta} \, \mathrm{d}\theta = 0, \tag{11.11}$$

and so from equation (11 10) a necessary and sufficient solvability condition is

$$\frac{\partial \Theta}{\partial T} \int_0^{2\pi} \left(\frac{\partial u_0}{\partial \theta} \right)^2 \mathrm{d}\theta = \int_0^{2\pi} \frac{\partial u_0}{\partial \theta} \mathcal{L}_1 u_0 \, \mathrm{d}\theta \tag{11 12}$$

Owing to rotation and translation symmetry, we know that

$$\int_0^{2\pi} \left(\frac{\partial u_0}{\partial \theta} \right)^2 \mathrm{d}\theta \tag{11.13}$$

will depend only on k^2 Now $\mathcal{L}_1$ will contain powers of

$$\nabla_x = \boldsymbol{k} \frac{\partial}{\partial \theta} + \epsilon \nabla_X. \tag{11.14}$$

The large-scale spatial derivatives ∇_X of u_0 are deduced from

$$\nabla_X u_0 = \frac{\partial u_0}{\partial k^2} \nabla_X k^2, \tag{11 15}$$

since all the large-scale variation is to be found in the wavevector, $\boldsymbol{k}$, and u_0 can only depend on the wavevector through k^2, owing to rotation symmetry. Thus $\mathcal{L}_1 u_0$ will contain terms of the form $\nabla_X \quad \boldsymbol{k}$ and $\boldsymbol{k} \quad \nabla_X k^2$ After the θ dependence on the righthand side of equation (11 12) is integrated out, the phase equation (11.12) is therefore left in the form

$$\frac{\partial \Theta}{\partial T} = k D_\perp(k) \nabla \quad \widehat{\boldsymbol{k}} + D_\parallel(k) (\widehat{\boldsymbol{k}} \cdot \nabla) k, \tag{11.16}$$

where $k = |\boldsymbol{k}|$, $\widehat{\boldsymbol{k}} = \boldsymbol{k}/k$, and $\nabla \equiv \nabla_X$. The functions $D_\perp(k)$ and $D_\parallel(k)$ are analytic by construction

Now, we knew from the beginning that the system was homogeneous and isotropic, and so we could have inferred that the phase equation for $\partial \Theta / \partial I$ would depend only on k^2 and space derivatives of $\widehat{\boldsymbol{k}}$ Then clearly equation (11 16) is the most general form that the phase equation could take at leading order in the space derivatives

The reason for naming the functions $D_\perp(k)$ and $D_\parallel(k)$ can be seen by substituting $\boldsymbol{k} = \nabla_X \Theta$ to get

$$\frac{\partial \Theta}{\partial T} = D_\perp(k) \nabla^2 \Theta + (D_\parallel(k) - D_\perp(k)) \frac{k_i k_j}{k^2} \frac{\partial^2 \Theta}{\partial X_i \partial X_j} \tag{11 17}$$

Now in the case where the rolls are almost parallel, we may define the Cartesian x direction via $\boldsymbol{k} = k\hat{\boldsymbol{x}}$, where $\hat{\boldsymbol{x}}$ is a unit vector in the x direction. Defining the Cartesian y direction accordingly, the phase equation reduces to

$$\frac{\partial \Theta}{\partial T} = D_{\parallel}(k)\frac{\partial^2 \Theta}{\partial X^2} + D_{\perp}(k)\frac{\partial^2 \Theta}{\partial Y^2} \tag{11.18}$$

We can see now that $D_{\parallel}(k)$ and $D_{\perp}(k)$ are the diffusion coefficients for disturbances parallel and perpendicular to the wavevector respectively.

Equation (11.16) can be rewritten in the form

$$\tau(k)\frac{\partial \Theta}{\partial T} + \nabla \quad (\boldsymbol{k}B(k)) = 0, \tag{11.19}$$

subject to certain conditions on $D_{\parallel}(k)$ and $D_{\perp}(k)$, discussed below, which ensure that $B(k)$ is single-valued. This is the **Cross–Newell equation**. It will be shown below that for variational systems $\tau(k)$ is positive for all k, and we shall assume that for a general system $\tau(k)$ does not change sign for k in the band of interest, so that we may choose $\tau(k) > 0$ by assigning the sign of $B(k)$ appropriately.

Expanding out the second term in equation (11.19) gives

$$\tau(k)\frac{\partial \Theta}{\partial T} + kB(k)\nabla \cdot \hat{\boldsymbol{k}} + \left(B(k) + k\frac{\mathrm{d}B}{\mathrm{d}k}\right)(\hat{\boldsymbol{k}} \cdot \nabla)k = 0, \tag{11.20}$$

and so comparing with equation (11.16) we have

$$D_{\perp}(k) = -\frac{1}{\tau(k)}B(k), \tag{11.21}$$

$$D_{\parallel}(k) = -\frac{1}{\tau(k)}\frac{\mathrm{d}(kB)}{\mathrm{d}k}. \tag{11.22}$$

The phase equation (11.20) can be rewritten in terms of spatial derivatives of the phase by setting $\boldsymbol{k} = (k_x, k_y)$ to arrive at

$$\tau(k)\frac{\partial \Theta}{\partial T} = \left(B + \frac{k_x^2}{k}\frac{\mathrm{d}B}{\mathrm{d}k}\right)\frac{\partial^2 \Theta}{\partial X^2} + \frac{2k_x k_y}{k}\frac{\mathrm{d}B}{\mathrm{d}k}\frac{\partial^2 \Theta}{\partial X \partial Y} + \left(B + \frac{k_y^2}{k}\frac{\mathrm{d}B}{\mathrm{d}k}\right)\frac{\partial^2 \Theta}{\partial Y^2} \tag{11.23}$$

If c_{XX} is the coefficient of $\partial^2\Theta/\partial X^2$ in the equation for $\partial\Theta/\partial T$ and so on, then the equation is defined to be elliptic if $c_{XY}^2 < 4c_{XX}c_{YY}$ and hyperbolic if $c_{XY}^2 > 4c_{XX}c_{YY}$. The phase equation is therefore elliptic if $B\mathrm{d}(kB)/\mathrm{d}k > 0$ and hyperbolic if $B\mathrm{d}(kB)/\mathrm{d}k < 0$.

Looking at expressions (11.21) and (11.22) for the diffusion coefficients, we can see that $D_{\parallel}(k)$ and $D_{\perp}(k)$ will be positive, and thus the rolls will be stable if $B(k)/\tau(k) < 0$ and $(1/\tau(k))\mathrm{d}(kB)/\mathrm{d}k < 0$. Hence we need $B(k) < 0$ and

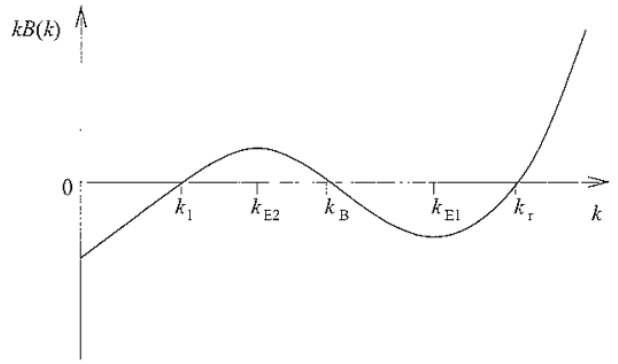


Fig 11 1. The typical shape of the function $kB(k)$ The Busse stability balloon is bounded by $k = k_B$ and $k = k_{E1}$, the zigzag and Eckhaus stability boundaries respectively The neutral stability boundaries are at $k = k_1$ and $k = k_1$ At $k = k_{E2}$ the mode of instability changes from pure zigzag to mixed

$d(kB)/dk < 0$ for stability If $B(k) < 0$ but $d(kB)/dk > 0$ then disturbances with spatial variation in the direction parallel to the wavevector can grow, perpendicular disturbances are suppressed, and the Eckhaus instability is seen. Conversely if $d(kB)/dk < 0$, but $B(k) > 0$ the growing disturbances are those perpendicular to the wavevector and the zigzag instability manifests itself When both $d(kB)/dk$ and $B(k)$ are positive, perturbations with components in both directions will grow.

The typical shape of the function $kB(k)$ is shown in Figure 11.1 Note that inside the Busse stability balloon, $k_B \leq k \leq k_{E1}$, the Cross–Newell equation is elliptic-stable. As we move outside the Busse balloon the equation becomes hyperbolic and is ill-posed because either the perpendicular or the parallel diffusion coefficient is negative In any natural pattern it is likely that the local wavevector will fall outside the Busse balloon in certain places from time to time, so this ill-posedness is of some concern For $k < k_B$, rolls are unstable to the zigzag instability, which is supercritical and so the instability saturates with the local wavenumber close to the original unstable one. In this case the phase equation (11 20) is easily regularised by the addition on the lefthand side of a term

$$\epsilon^2 \frac{1}{4k_B} \left| \frac{dB(k_B)}{dk} \right| \nabla^2 \nabla \cdot \boldsymbol{k}, \tag{11 24}$$

which is the most important term arising at the next order of the perturbation expansion. However when the local wavenumber lies in the Eckhaus unstable region $k > k_{E1}$, a problem arises because the instability is subcritical and so leads to a large change in the local wavenumber. In fact for the Eckhaus instability the final local wavenumber may even lie outside the marginal stability curve, so we

can no longer assume that the pattern is fully nonlinear with an amplitude that is not close to zero and is slaved to the phase. The regularisation of the phase equation in this case requires the amplitude to be reincorporated as an active parameter. Passot and Newell dealt with this issue by introducing an equation for the order parameter $U = A\cos\theta$, where A is the local amplitude of the rolls. This order parameter equation is simpler than the original governing partial differential equations from which it is derived and reduces to the Cross–Newell equation inside the Busse balloon, but crucially describes the behaviour of the pattern when the local wavenumber lies outside. We shall not consider this order parameter equation any further here, but details can be found in Passot and Newell (1994). Amplitude regularisation is typically also needed in the vicinity of foci, where the high curvature renders the phase description invalid (see Newell, Passot & Souli, 1990).

At the zigzag stability boundary, $k = k_B$, the perpendicular diffusion coefficient $D_\perp(k)$ is zero, and therefore the function $D_\parallel(k)/kD_\perp(k)$ must have a pole there. Combining equations (11.21) and (11.22) gives

$$\frac{\mathrm{d}(\ln kB(k))}{\mathrm{d}k} = \frac{D_\parallel(k)}{kD_\perp(k)} \tag{11.25}$$

If we integrate this round a closed curve Γ in wavevector space enclosing the pole at $k = k_B$, then since $D_\parallel(k)$ and $D_\perp(k)$ are analytic, we find that the change in the logarithm around the curve is given by

$$[\ln(kB(k))]_\Gamma = 2\pi i r, \tag{11.26}$$

where the residue r at the pole $k = k_B$ is given by

$$\begin{aligned} r &= \lim_{k\to k_B} \frac{(k-k_B)D_\parallel(k)}{kD_\perp(k)} \\ &= \lim_{k\to k_B} \frac{\mathrm{d}((k-k_B)D_\parallel(k))/\mathrm{d}k}{\mathrm{d}(kD_\perp(k))/\mathrm{d}k} \text{ (by l'Hôpital's rule)} \\ &= \lim_{k\to k_B} \frac{D_\parallel(k) + (k-k_B)\mathrm{d}(D_\parallel(k))/\mathrm{d}k}{k\mathrm{d}(D_\perp(k))/\mathrm{d}k + D_\perp(k)} \\ &= \frac{D_\parallel(k_B)}{k_B\mathrm{d}(D_\perp(k_B))/\mathrm{d}k}. \end{aligned} \tag{11.27}$$

For $kB(k)$ to be single-valued, the change in its logarithm around a closed curve in wavevector space must be $2\pi i$, hence we must have $r = 1$ if $kB(k)$ is to be analytic. In that case we can integrate equation (11.25) to get

$$kB(k) = k_0 B(k_0) \exp \int_{k_0}^{k} \frac{D_\parallel(k)}{kD_\perp(k)}\mathrm{d}k, \tag{11.28}$$

for some constant reference wavenumber, k_0. In order to write the Cross–Newell equation in the form of equation (11.19) we must therefore satisfy $r = 1$. If we can do this, then we can identify a formal free energy

$$\overline{\mathcal{F}} = \int_{\mathcal{D}} \left(-\frac{1}{2} \int^{k^2} \widetilde{B}(z) \mathrm{d}z \right) \mathrm{d}X\mathrm{d}Y, \tag{11.29}$$

where $\widetilde{B}(k^2) \equiv B(k)$ and where $\mathcal{D}$ is the domain of pattern formation, since then

$$\begin{aligned} \delta\overline{\mathcal{F}} &= \int_{\mathcal{D}} \left(-\frac{1}{2} \widetilde{B}(k^2) \delta(k^2) \right) \mathrm{d}X\mathrm{d}Y, \\ &= \int_{\mathcal{D}} \left(-\frac{1}{2} \widetilde{B}(k^2) \delta(\nabla\Theta \cdot \nabla\Theta) \right) \mathrm{d}X\mathrm{d}Y, \\ &= \int_{\mathcal{D}} \left(-\widetilde{B}(k^2) \nabla\Theta \quad \nabla\delta\Theta \right) \mathrm{d}X\mathrm{d}Y, \\ &= \int_{\mathcal{D}} \left(-\widetilde{B}(k^2) \boldsymbol{k} \quad \nabla\delta\Theta \right) \mathrm{d}X\mathrm{d}Y, \end{aligned} \tag{11.30}$$

which upon application of the divergence theorem becomes

$$\delta\overline{\mathcal{F}} = -\int_{\mathcal{C}} B(k) \boldsymbol{k} \quad \widehat{\boldsymbol{s}}\, \delta\Theta \, \mathrm{d}s + \int_{\mathcal{D}} \nabla \quad (\boldsymbol{k} B(k))\, \delta\Theta \, \mathrm{d}X\mathrm{d}Y, \tag{11.31}$$

where $\mathcal{C}$ is the system boundary, $\widehat{\boldsymbol{s}}$ is the outward normal on the boundary and s is the arc length along the boundary. Now if we have periodic boundary conditions and no singularities of the phase field within $\mathcal{C}$ then the integral around $\mathcal{C}$ vanishes. This is of course highly unrealistic, as there will almost certainly be singularities somewhere in the domain. Another possibility for getting the first integral to vanish is to require that $\boldsymbol{k} \cdot \widehat{\boldsymbol{s}}$ is zero everywhere on the boundary. This is not so unrealistic: it just says that rolls are locally normal to the boundary, which is in fact the energetically preferred configuration in Rayleigh–Bénard convection (see Chapter 10). Finally, we could also insist that the boundary passes only through regions where the pattern adopts the critical wavenumber so that B is zero and again the boundary integral vanishes. If we have succeeded in getting rid of the first integral then we can observe finally that

$$\frac{\delta\overline{\mathcal{F}}}{\delta\Theta} = \nabla \cdot (\boldsymbol{k} B(k)) = -\tau(k) \frac{\partial\Theta}{\partial T}, \tag{11.32}$$

and

$$\frac{\mathrm{d}\overline{\mathcal{F}}}{\mathrm{d}T} = -\int_{\mathcal{D}} \tau(k) \left(\frac{\partial\Theta}{\partial T} \right)^2 \mathrm{d}X\mathrm{d}Y \leq 0 \tag{11.33}$$

So $\overline{\mathcal{F}}$ is always decreasing, and hence $\overline{\mathcal{F}}$ is indeed a Lyapunov functional or free energy. This is the case even if the original governing equation (11.1) cannot be put into gradient form and has no free energy.

In fact we have cheated a little bit in our consideration of the boundary integral. The boundary curve, $\mathcal{C}$, must include not only the obvious boundaries of the system, but also curves around any singularities of the phase variable, such as dislocations and foci. To deal with these we put a contour of length $O(\epsilon)$ around each singularity. The Cross–Newell theory no longer holds at such short lengthscales, but if $\partial\Theta/\partial I$ remains $O(1)$ anyway, then provided there are fewer than $O(\epsilon^{-1})$ singularities, $\mathrm{d}\overline{\mathcal{F}}/\mathrm{d}T$ is still negative at leading order and $\overline{\mathcal{F}}$ is still decreasing. In their original paper Cross and Newell (1984) argued that the singularities in the pattern are likely to be isolated foci, and that $\overline{\mathcal{F}}$ will act as a Lyapunov functional over patches centred on each focus, while a higher order equation is needed to match between the patches. If on the other hand moving dislocations are present, then $\overline{\mathcal{F}}$ may increase, except in special circumstances.

11.2 Stationary solutions of the Cross–Newell equation

If $\overline{\mathcal{F}}$ is in fact always decreasing, and is bounded below, then eventually $\overline{\mathcal{F}}$ will reach its minimum value and we will have $\mathrm{d}\overline{\mathcal{F}}/\mathrm{d}T = 0$. From equation (11.33) we then see that $\partial\Theta/\partial I$ must be zero everywhere in the system, since $\tau(k)$ is always positive. Now since $B(k)$ is analytic by construction if the residue satisfies $r = 1$, and since we have $k_1 < k < k_r$ and a finite system size, $\overline{\mathcal{F}}$ must indeed be bounded below. This means that if $\overline{\mathcal{F}}$ is indeed acting as a free energy the pattern will eventually settle down to a stationary solution of the Cross–Newell equation.

Stationary solutions of the phase equation (11.19) must satisfy

$$\nabla \times \boldsymbol{k} = \boldsymbol{0}, \tag{11.34}$$

$$\nabla \quad (\boldsymbol{k}B(k)) = 0 \tag{11.35}$$

If we transform to new coordinates $\alpha(X, Y) = \Theta$ and $\beta(X, Y)$ defined by

$$\frac{\partial \alpha}{\partial X} = k\cos\psi, \tag{11.36}$$

$$\frac{\partial \alpha}{\partial Y} = k\sin\psi, \tag{11.37}$$

$$\frac{\partial \beta}{\partial X} = -l\sin\psi, \tag{11.38}$$

$$\frac{\partial \beta}{\partial Y} = l\cos\psi, \tag{11.39}$$

where $\boldsymbol{k} = (k\cos\psi, k\sin\psi)$ and l is a measure of local roll curvature, then equation (11 35) becomes

$$k^2\frac{\partial B}{\partial\alpha} + Bk\frac{\partial k}{\partial\alpha} + Bkl\frac{\partial\psi}{\partial\beta} = 0 \tag{11.40}$$

The compatibility condition $\partial^2\beta/\partial X\partial Y = \partial^2\beta/\partial Y\partial X$ gives

$$\frac{\partial}{\partial Y}(-l\sin\psi) = \frac{\partial}{\partial X}(l\cos\psi) \tag{11 41}$$

and hence

$$\left(k\sin\psi\frac{\partial}{\partial\alpha} + l\cos\psi\frac{\partial}{\partial\beta}\right)(-l\sin\psi) = \left(k\cos\psi\frac{\partial}{\partial\alpha} - l\sin\psi\frac{\partial}{\partial\beta}\right)(l\cos\psi) \tag{11 42}$$

Expanding out all the terms tells us eventually that

$$\frac{\partial\psi}{\partial\beta} = -\frac{k}{l^2}\frac{\partial l}{\partial\alpha}, \tag{11.43}$$

which when substituted into equation (11.40) gives

$$kl\frac{\partial}{\partial\alpha}\left(\frac{kB(k)}{l}\right) = 0. \tag{11 44}$$

Stationary solutions of the phase equation must therefore satisfy

$$\frac{kB(k)}{l} = H(\beta), \tag{11.45}$$

for some arbitrary function $H(\beta)$.

Now curves of constant α are the contours of constant phase, and so the α coordinate measures distance in a direction parallel to the local wavevector and perpendicular to the roll axes. The curves of constant β are orthogonal to the phase contours, and the β coordinate measures distance along the roll axes. So $kB(k)/l$ is constant along trajectories perpendicular to the roll axes.

In the vicinity of a focus singularity l becomes large, as it is a measure of the curvature of the rolls, and reaches a magnitude $O(\epsilon^{-1})$ where the Cross–Newell equations break down. This fixes $H(\beta)$ to be $O(\epsilon)$. If we now move away from the centre of the focus on any line of constant β, we retreat into a region free of singularities and once again we have $O(1)$ values of l. So now we have

$$kB(k) = c(\beta)l\epsilon, \tag{11.46}$$

where $H(\beta) = \epsilon c(\beta)$, and hence

$$k = k_0 + O(\epsilon), \tag{11 47}$$

where $B(k_0) = 0$. So in any region that is connected to the immediate neighbourhood of a focus singularity by a trajectory orthogonal to the rolls, we have wavenumber selection. The selected wavenumber k_0 must lie within the Busse balloon if the Cross–Newell equation is to hold, but since $B(k_0)$ is zero we must have $k_0 = k_B$ and so it also lies at the zigzag instability boundary.

11.3 Defect solutions of the Cross–Newell equation

The Cross–Newell formalism provides a framework for the study of pattern defects in systems far from equilibrium. Typical defects in a roll pattern close to onset are dislocations and amplitude grain boundaries, where the pattern amplitude tends to zero (see Chapter 10). Far from onset a wider variety of defects can be seen and the amplitude need not be zero at the core.

11.3.1 Phase grain boundaries

In the hyperbolic region, $k_{E2} < k < k_B$, the stationary phase equations (11.34) and (11.35) have weak or shock solutions which define phase grain boundaries, where the wavevector changes direction discontinuously, but the amplitude remains nonzero (Figure 11.2). The biharmonic term (11.24) regularises the shocks, but in fact the exact form of the regularisation is unimportant as long as it is in flux-divergence form. The standard theory of characteristics can be applied and is given in detail in Passot and Newell (1994); here we will sketch just one important result.

Starting from the stationary regularised phase equation

$$\nabla \quad (\boldsymbol{k}B(k)) + \epsilon^2 \frac{1}{4k_B} \left| \frac{dB(k_B)}{dk} \right| \nabla^2 \nabla \cdot \boldsymbol{k} = 0, \tag{11.48}$$

we seek solutions with discontinuities along shock lines

$$\frac{dX}{dY} = s = \cot\phi, \tag{11.49}$$

where s is a constant that defines the slope of the shock, and ϕ is the angle that the shock makes with the X axis. Solutions of equation (11.48) that depend only on $Z = X - sY$, where Z measures distance along a line perpendicular to the shock, must satisfy

$$\frac{\partial}{\partial Z}\left(kB(k)\cos\psi + \epsilon^2 \frac{1}{4k_B}\left|\frac{dB(k_B)}{dk}\right| \frac{\partial(\nabla \quad \boldsymbol{k})}{\partial Z}\right)$$
$$-s\frac{\partial}{\partial Z}\left(kB(k)\sin\psi - s\epsilon^2 \frac{1}{4k_B}\left|\frac{dB(k_B)}{dk}\right| \frac{\partial(\nabla \quad \boldsymbol{k})}{\partial Z}\right) = 0, \tag{11.50}$$

where $\boldsymbol{k} = (k\cos\psi, k\sin\psi)$ Integrating this equation with respect to Z between $\pm\infty$, gives

$$\left[kB(k)\cos\psi + \epsilon^2\frac{1}{4k_B}\left|\frac{dB(k_B)}{dk}\right|\frac{\partial(\nabla\cdot\boldsymbol{k})}{\partial Z}\right]_{-\infty}^{+\infty}$$
$$-s\left[kB(k)\sin\psi - s\epsilon^2\frac{1}{4k_B}\left|\frac{dB(k_B)}{dk}\right|\frac{\partial(\nabla\cdot\boldsymbol{k})}{\partial Z}\right]_{-\infty}^{+\infty} = 0. \qquad (11\,51)$$

where $[\,]_{-\infty}^{+\infty}$ represents the difference between the value of a quantity as $Z\to+\infty$ and its value as $Z\to-\infty$ In the limit $\epsilon\to 0$, this becomes the difference across the shock at $Z=0$, and gives the shock condition

$$[k\cos\psi B(k)]_{0-}^{0+} - s[k\sin\psi B(k)]_{0-}^{0+} = 0 \qquad (11\,52)$$

Obviously the exact form of the regularisation is not important as long as it is in flux-divergence form, $\nabla\ (F, G)$, where F and G tend to zero as $Z\to\pm\infty$

A second jump or shock condition is imposed by the equation

$$\nabla\times\boldsymbol{k} = \boldsymbol{0} \qquad (11\,53)$$

For solutions that depend only on Z, this becomes

$$\frac{\partial}{\partial Z}(k\sin\psi + sk\cos\psi) = 0, \qquad (11\,54)$$

and hence the jump condition is

$$[k\sin\psi + sk\cos\psi]_{0-}^{0+} = 0. \qquad (11\,55)$$

If the wavenumber k is continuous on each side of the shock or phase grain boundary, the shock conditions reduce to

$$[\sin\psi + s\cos\psi]_{0-}^{0+} = 0, \qquad (11.56)$$

$$[\cos\psi B(k) - s\sin\psi B(k)]_{0-}^{0+} = 0 \qquad (11\,57)$$

If $B(k)\neq 0$, then according to these equations we must have $[\sin\psi]_{0-}^{0+} = [\cos\psi]_{0-}^{0+} = 0$ and so there is no discontinuity in the wavevector across the grain boundary So we must choose $B(k)=0$, which means that the wavenumber must lie at the zigzag boundary $k=k_B$ In this case we must satisfy only

$$[\sin\psi + s\cos\psi]_{0-}^{0+} = 0 \qquad (11.58)$$

Choosing $\psi=\pi/2$ at $Z=0_-$, implies that at $Z=0_+$ we must have $\psi\equiv\psi_+$, where

$$\sin\psi_+ + s\cos\psi_+ = 1. \qquad (11\,59)$$

Since $\sin^2\psi_+ + \cos^2\psi_+ = 1$, ψ_+ must satisfy

$$\cos^2\psi_+ + (1 - s\cos\psi_+)^2 = 1, \tag{11.60}$$

and so

$$\cos\psi_+((1+s^2)\cos\psi_+ - 2s) = 0. \tag{11.61}$$

Either $\cos\psi_+ = 0$, and so from equation (11.59) $\sin\psi_+ = 1$, so that $\psi_+ = \pi/2$ and there is no discontinuity across the shock, or ψ_+ is given by

$$\cos\psi_+ = \frac{2s}{1+s^2}, \tag{11.62}$$

$$\sin\psi_+ = \frac{1-s^2}{1+s^2}. \tag{11.63}$$

Since $s = \cot\phi$, where ϕ is the angle that the characteristic makes with the X axis, we can write

$$\cos\psi_+ = \sin 2\phi, \tag{11.64}$$

$$\sin\psi_+ = -\cos 2\phi, \tag{11.65}$$

and we can deduce that $\psi_+ = 2\phi - \pi/2$. Now the contours of the pattern are given by level lines of the phase, which are locally orthogonal to the wavevector. So on the $Z = 0_-$ side of the shock the roll contours lie parallel to the X axis and on the $Z = 0_+$ side of the shock they lie at an angle 2ϕ to the X axis. The shock angle is the average of the angles of the roll contours on either side of the shock, and the shock, or phase grain boundary, bisects the wavevectors, as shown in Figure 11.2.

11.3.2 Point defects

The characteristic point defect of a roll pattern close to onset is the dislocation. Far from onset, **disclinations** are the basic point defects from which all others can be constructed. This fundamental difference arises because the phase is no longer single-valued far from onset.

To describe rolls close to onset, we write $u = Ae^{i\theta} + c.c.$ at leading order, where $\theta = \boldsymbol{k}\cdot\boldsymbol{x}$, and follow the evolution of the complex envelope A using the Newell–Whitehead–Segel equation, which takes variational form. The phase, θ, is uniquely defined in this case, and consequently so is the wavevector, $\boldsymbol{k}$. Far from onset, however, we no longer have a single pair of wavevectors, $\pm\boldsymbol{k}$, contributing to the pattern at leading order and we must write $u = A\cos\theta$, where A is real. The phase, θ, is now double-valued, since the transformation $\theta \to -\theta$ leaves A and

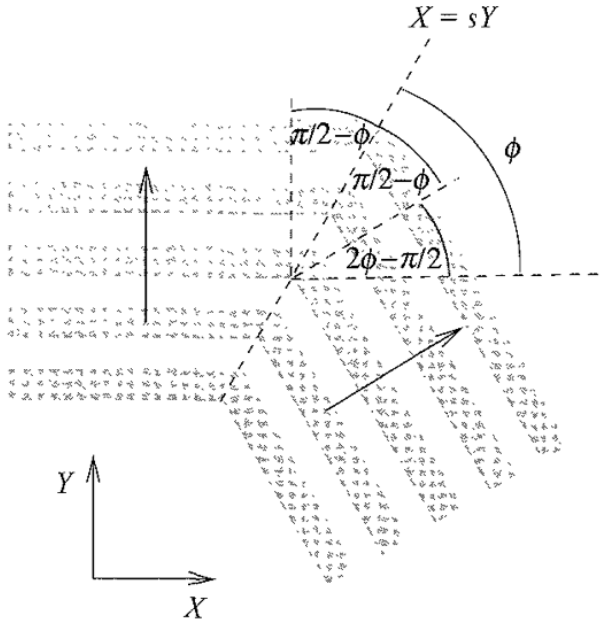


Fig. 11.2. Rolls and their associated wavevectors (shown as arrows) on either side of a phase grain boundary. The grain boundary, $X = sY$, lies at an angle $\phi = \cot^{-1}(s)$ to the X axis, bisecting the wavevectors on either side

u unchanged. (Note that in the complex, close to onset, case this transformation would correspond to $A \to \bar{A}$.) The wavevector, $\boldsymbol{k} = \nabla\theta$, is not unique either, since $\boldsymbol{k}$ and $-\boldsymbol{k}$ cannot be distinguished. Far from onset, only the director field in which $\pm\boldsymbol{k}$ are regarded as identical can be uniquely defined. This allows the development of disclination defects where the wavevector rotates through an angle $\pm\pi$ around the defect core (Figure 11.3).

A whole family of point defects, built from disclinations, are found for patterns far from equilibrium using the Cross–Newell formalism (Passot & Newell 1994). We first define a conjugate phase, $\widehat{\Theta}(k, \psi)$, such that

$$\Theta(X, Y) + \widehat{\Theta}(k, \psi) = \boldsymbol{k} \quad \boldsymbol{X}, \tag{11.66}$$

where $\boldsymbol{X} = (X, Y)$ and $\boldsymbol{k} = (k\cos\psi, k\sin\psi)$. Now for stationary solutions satisfying equations (11.34) and (11.35), X and Y will be functions of k and ψ and vice versa. Differentiating equation (11.66) with respect to k thus gives

$$\frac{\partial\Theta}{\partial X}\frac{\partial X}{\partial k} + \frac{\partial\Theta}{\partial Y}\frac{\partial Y}{\partial k} + \frac{\partial\widehat{\Theta}}{\partial k} = k\cos\psi\frac{\partial X}{\partial k} + k\sin\psi\frac{\partial Y}{\partial k} + X\cos\psi + Y\sin\psi \tag{11.67}$$

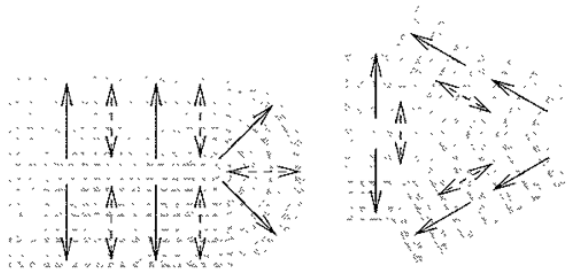


Fig. 11.3. Phase contours of the convex and concave disclinations (a) The wavevector, $\boldsymbol{k}$ (solid arrows), rotates through an angle π around the core of this convex dislocation and so is discontinuous across a line extending out from the defect core, while the director field, $\pm\boldsymbol{k}$ (dashed double-headed arrows), is continuous there (b) For the concave disclination the rotation is through $-\pi$

Using $\boldsymbol{k} = \nabla\Theta$ this reduces to

$$\frac{\partial\widehat{\Theta}}{\partial k} = X\cos\psi + Y\sin\psi \quad \left(= \frac{1}{k}\boldsymbol{k} \quad \boldsymbol{X}\right). \tag{11.68}$$

Similarly, differentiating equation (11.66) with respect to ψ gives

$$\frac{\partial\widehat{\Theta}}{\partial\psi} = -Xk\sin\psi + Yk\cos\psi \tag{11.69}$$

From equations (11.68) and (11.69) it is straightforward to deduce that

$$X = \cos\psi\frac{\partial\widehat{\Theta}}{\partial k} - \frac{\sin\psi}{k}\frac{\partial\widehat{\Theta}}{\partial\psi}, \tag{11.70}$$

$$Y = \sin\psi\frac{\partial\widehat{\Theta}}{\partial k} + \frac{\cos\psi}{k}\frac{\partial\widehat{\Theta}}{\partial\psi}, \tag{11.71}$$

and hence it is possible to work out, after a rather long and boring calculation, that

$$\frac{\partial}{\partial X} = \frac{1}{Det(k,\psi)}\left(a(k,\psi)\frac{\partial}{\partial k} + b(k,\psi)\frac{\partial}{\partial\psi}\right), \tag{11.72}$$

$$\frac{\partial}{\partial Y} = \frac{1}{Det(k,\psi)}\left(c(k,\psi)\frac{\partial}{\partial k} + d(k,\psi)\frac{\partial}{\partial\psi}\right), \tag{11.73}$$

where

$$a(k,\psi) = \cos\psi\frac{\partial\widehat{\Theta}}{\partial k} + \sin\psi\frac{\partial^2\widehat{\Theta}}{\partial\psi\partial k} - \frac{\sin\psi}{k}\frac{\partial\widehat{\Theta}}{\partial\psi} + \frac{\cos\psi}{k}\frac{\partial^2\widehat{\Theta}}{\partial\psi^2}, \tag{11.74}$$

$$b(k,\psi) = -\sin\psi\frac{\partial^2\widehat{\Theta}}{\partial k^2} + \frac{\cos\psi}{k^2}\frac{\partial\widehat{\Theta}}{\partial\psi} - \frac{\cos\psi}{k}\frac{\partial^2\widehat{\Theta}}{\partial\psi\partial k}, \tag{11.75}$$

$$c(k,\psi) = \sin\psi\frac{\partial\widehat{\Theta}}{\partial k} - \cos\psi\frac{\partial^2\widehat{\Theta}}{\partial\psi\partial k} + \frac{\cos\psi}{k}\frac{\partial\widehat{\Theta}}{\partial\psi} + \frac{\sin\psi}{k}\frac{\partial^2\widehat{\Theta}}{\partial\psi^2}, \tag{11.76}$$

$$d(k,\psi) = \cos\psi\frac{\partial^2\widehat{\Theta}}{\partial k^2} + \frac{\sin\psi}{k^2}\frac{\partial\widehat{\Theta}}{\partial\psi} - \frac{\sin\psi}{k}\frac{\partial^2\widehat{\Theta}}{\partial\psi\partial k}, \tag{11.77}$$

$$Det(k,\psi) = a(k,\psi)d(k,\psi) - b(k,\psi)c(k,\psi) \tag{11.78}$$

Substituting from equations (11.72) and (11.73) into equation (11.35) gives, eventually,

$$k\frac{\partial}{\partial k}\left(kB(k)\frac{\partial\widehat{\Theta}}{\partial k}\right) + \frac{\partial}{\partial k}(kB(k))\frac{\partial^2\widehat{\Theta}}{\partial\psi^2} = 0. \tag{11.79}$$

This equation is linear and separable, so we look for solutions of the form $\widehat{\Theta}(k,\psi) = K(k)\Psi(\psi)$, and find

$$\frac{k}{K}\left(\frac{\mathrm{d}K}{\mathrm{d}k} + \frac{kB(k)}{\mathrm{d}/\mathrm{d}k(kB(k))}\frac{\mathrm{d}^2K}{\mathrm{d}k^2}\right) = -\frac{1}{\Psi}\frac{\mathrm{d}^2\Psi}{\mathrm{d}\psi^2} = \omega^2, \tag{11.80}$$

where ω^2 is a constant. We choose ω^2 to be real, so that Ψ will be periodic in ψ, and we also choose $\omega = n$, where n is an integer, so that Ψ will be single-valued. Hence we have

$$\Psi(\psi) = \Psi_n(\psi) \equiv g_n\cos n(\psi - \psi_n), \tag{11.81}$$

where n is a positive integer and where g_n and ψ_n are real constants. The general solution for $n = 0$ is

$$\Psi(\psi) = \Psi_0(\psi) \equiv g_{00} + g_{01}\psi, \tag{11.82}$$

where g_{00} and g_{01} are real constants.

Now $K(k)$ must satisfy

$$\frac{\mathrm{d}^2K}{\mathrm{d}k^2} + \frac{1}{kB(k)}\frac{\mathrm{d}}{\mathrm{d}k}(kB(k))\frac{\mathrm{d}K}{\mathrm{d}k} - K\frac{n^2}{k^2B(k)}\frac{\mathrm{d}}{\mathrm{d}k}(kB(k)) = 0 \tag{11.83}$$

For $n = 0$ the solution is

$$K(k) = K_0k \equiv h_{00} + h_{01}\int^k\frac{\mathrm{d}k}{kB(k)}, \tag{11.84}$$

where h_{00} and h_{01} are real constants. The general solution for $\widehat{\Theta}$ is therefore

$$\widehat{\Theta} = s_0 + s_1\psi + (s_2 + s_3\psi)\int^k \frac{\mathrm{d}k}{kB(k)}, \tag{11.85}$$

where s_0, s_1, s_2 and s_3 are real constants. This corresponds to

$$X = \frac{1}{kB(k)}(s_2 + s_3\psi)\cos\psi - \frac{\sin\psi}{k}\left(s_1 + s_3\int^k \frac{\mathrm{d}k}{kB(k)}\right), \tag{11.86}$$

$$Y = \frac{1}{kB(k)}(s_2 + s_3\psi)\sin\psi + \frac{\cos\psi}{k}\left(s_1 + s_3\int^k \frac{\mathrm{d}k}{kB(k)}\right), \tag{11.87}$$

$$\Theta = -(s_0 + s_1\psi) + (s_2 + s_3\psi)\left(\frac{1}{B(k)} - \int^k \frac{\mathrm{d}k}{kB(k)}\right) \tag{11.88}$$

For $s_1 = s_3 = 0$, the phase, Θ, is purely a function of k, so k is constant on contours of constant phase. We also have

$$X^2 + Y^2 = \frac{s_2^2}{k^2B(k)^2}, \tag{11.89}$$

and so $X^2 + Y^2$ is constant on the phase contours, which must therefore be circles. This defect is therefore a target pattern, alternatively known as a focus. For $s_2 = s_3 = 0$, the phase is given by

$$\Theta = -s_0 - s_1\psi = -s_0 - s_1\tan^{-1}(X/Y), \tag{11.90}$$

and so X/Y is constant on contours of constant phase. This is a **vortex** solution (Figure 11.4a). A general solution for arbitrary s_i combines features of the focus and the vortex and so is a spiral. If we take the particular example where $s_3 = 0$, we find

$$\Theta = -s_0 - s_1\psi + s_2\left(\frac{1}{B(k)} - \int^k \frac{\mathrm{d}k}{kB(k)}\right), \tag{11.91}$$

$$R^2 = \frac{s_2^2}{k^2B(k)^2} + \frac{s_1^2}{k^2}, \tag{11.92}$$

$$\phi = \psi + \tan^{-1}(s_1B(k)/s_2), \tag{11.93}$$

where (R, ϕ) are polar coordinates for X and Y. Equation (11.92) says that k is a function of R only, and thus the phase can be rewritten

$$\Theta = -s_0 - s_1\phi + h(R), \tag{11.94}$$

for a function $h(R)$ that can be determined from equations (11 92) and (11 93). Comparing this with the definition (10.94), we can see that the solution is an m-armed spiral if $s_1 = m$, where m is an integer

For $n = 1$ we have

$$K(k) = K_1(k) \equiv h_{10}k + h_{11}k \int^k \frac{\mathrm{d}k}{k^3 B(k)}, \tag{11 95}$$

where h_{10} and h_{11} are real constants, and so the general solution for $\widehat{\Theta}$ is

$$\widehat{\Theta} = s_4 k \int^k \frac{\mathrm{d}k}{k^3 B(k)} \cos(\psi - \psi_{10}) + s_5 k \cos(\psi - \psi_{11}), \tag{11.96}$$

where s_4, s_5, ψ_{10} and ψ_{11} are real constants. If we fix the origin of ψ so that $\psi_{10} = 0$ we have

$$X = s_4 \int^k \frac{\mathrm{d}k}{k^3 B(k)} + \frac{s_4}{2k^2 B(k)}(1 + \cos 2\psi) + s_5 \cos \psi_{11}, \tag{11.97}$$

$$Y = \frac{s_4}{2k^2 B(k)} \sin 2\psi + s_5 \sin \psi_{11}, \tag{11 98}$$

$$\Theta = \frac{s_4}{kB(k)} \cos \psi. \tag{11 99}$$

To work out the shape of the phase contours first note that the free energy (11.29) is minimised when $\widetilde{B}(k^2) \equiv B(k) = 0$ everywhere in the pattern. We expect the wavenumber to lie in the stable region $k_B \leq k \leq k_{E1}$, and so if $B(k)$ is to be close to zero, we must have $k \approx k_B$. Thus the energetically preferred pattern has wavenumber k close to k_B. Now $B(k_B)$ is zero, so if $(1 + \cos 2\psi)$ is not close to zero, in other words if ψ is not close to $\pm\pi/2$, the dominant terms in equations (11.97) and (11 98) will be

$$X = \frac{s_4}{2k^2 B(k)}(1 + \cos 2\psi), \tag{11 100}$$

$$Y = \frac{s_4}{2k^2 B(k)} \sin 2\psi, \tag{11.101}$$

which gives

$$R^2 = \frac{s_4^2}{k^4 B(k)^2} \cos^2 \psi = \frac{\Theta^2}{k^2}, \tag{11 102}$$

$$\phi = \psi \text{ or } \psi + \pi. \tag{11 103}$$

Thus R will be constant on the phase contours if ϕ is not close to $\pm\pi/2$. On the other hand for $\psi \approx \pm\pi/2$, the dominant terms in the expressions for X and

Y will be

$$X = s_4 \int^k \frac{\mathrm{d}k}{k^3 B(k)}, \tag{11.104}$$

$$Y = \frac{s_4}{2k^2 B(k)} \sin 2\psi. \tag{11.105}$$

Expanding $B(k)$ in a Taylor series around $k = k_\mathrm{B}$ gives

$$X \approx s_4 \int^k \frac{\mathrm{d}k}{(k - k_\mathrm{B}) k_\mathrm{B}^3 \frac{dB(k_\mathrm{B})}{\mathrm{d}k}}, \tag{11.106}$$

and so

$$|k - k_\mathrm{B}| \approx \exp\left\{ \frac{k_\mathrm{B}^3 \frac{\mathrm{d}B(k_\mathrm{B})}{\mathrm{d}k}}{s_4} X \right\} \tag{11.107}$$

Recall from Figure 11.1 that $\mathrm{d}B(k_\mathrm{B})/\mathrm{d}k$ is negative, and so k approaches k_B exponentially fast as $X \to +\infty$ if $s_4 > 0$ or as $X \to -\infty$ if $s_4 < 0$. If $k \approx k_\mathrm{B}$ then from equation (11.99) the contours of constant Θ are approximately curves of constant ψ, and so from equation (11.105) they are approximately lines of constant Y. Thus in the right half-plane (for $s_4 > 0$) or left half-plane (for $s_4 < 0$) the phase contours look like straight lines parallel to the X axis, which join smoothly onto semicircles in the other half-plane, giving a roman arch shape. This is the convex disclination, sketched in Figure 11.3a for the case $s_4 < 0$.

For $n > 1$ there may not be an exact solution for $K(k)$, but we can make an approximation as follows. As above, we expect that the pattern wavenumber is close to k_B. Since $B(k_\mathrm{B}) = 0$, equation (11.83) has a regular singular point at $k = k_\mathrm{B}$. We therefore expand $B(k)$ in a Taylor series around $k = k_\mathrm{B}$ and $K(k)$ in a Frobenius expansion of the form

$$K(k) = h_{n0} \ln|k_\mathrm{B} - k|(1 + h_{n1}(k - k_\mathrm{B})) + h_{n2}(k - k_\mathrm{B}) + h\,o\,t\,, \tag{11.108}$$

where h_{n0}, h_{n1} and h_{n2} are real constants (see, for example, Arfken, 2001 for details of the Frobenius method). Substituting into equation (11.83) we find that $h_{n1} = n^2/k_\mathrm{B}$. It is also possible to determine h_{n2} in terms of the derivatives of $B(k)$ at $k = k_\mathrm{B}$, but it turns out that we don't need to know the exact form of h_2 here. Thus the solution for $\widehat{\Theta}$ is

$$\widehat{\Theta} = s_{n0} \left\{ \ln|k_\mathrm{B} - k| \left(1 + \frac{n^2}{k_\mathrm{B}}(k - k_\mathrm{B}) \right) + s_{n1}(k - k_\mathrm{B}) \right\} \cos n(\psi - \psi_n) + h.o.t., \tag{11.109}$$

where s_{n0}, s_{n1} and ψ_n are real constants. Using the freedom to choose the origin of ψ we can set ψ_n to zero, and then we have

$$X \approx s_{n0}\left(\frac{1}{k-k_B}+\frac{n^2}{k_B}\ln|k-k_B|\right)\cos\psi\cos n\psi + s_{n0}\frac{n}{k}\ln|k-k_B|\sin\psi\sin n\psi, \tag{11.110}$$

$$Y \approx s_{n0}\left(\frac{1}{k-k_B}+\frac{n^2}{k_B}\ln|k-k_B|\right)\sin\psi\cos n\psi - s_{n0}\frac{n}{k}\ln|k-k_B|\cos\psi\sin n\psi, \tag{11.111}$$

$$\Theta \approx s_{n0}\left\{\frac{k}{k-k_B}+\ln|k_B-k|\left(\frac{n^2k}{k_B}-1\right)\right\}\cos n\psi \tag{11.112}$$

When $\cos n\psi$ is not close to zero, we have

$$X \approx \frac{s_{n0}}{k-k_B}\cos\psi\cos n\psi, \tag{11.113}$$

$$Y \approx \frac{s_{n0}}{k-k_B}\sin\psi\cos n\psi, \tag{11.114}$$

$$\Theta \approx \frac{s_{n0}k}{k-k_B}\cos n\psi \tag{11.115}$$

to leading order, and so $R^2 \approx \Theta^2/k^2$. Thus away from $\cos n\psi = 0$ the phase contours are approximately arcs of circles. Close to $\cos n\psi = 0$, on the other hand, the logarithmic terms dominate in X and Y, and we have

$$X \approx s_{n0}\frac{n}{k}\ln|k-k_B|\sin\psi\sin n\psi, \tag{11.116}$$

$$Y \approx -s_{n0}\frac{n}{k}\ln|k-k_B|\cos\psi\sin n\psi. \tag{11.117}$$

From these last two equations we deduce that

$$|k-k_B| \approx \exp\left\{\frac{k(X\sin\psi - Y\cos\psi)}{s_{n0}n\sin n\psi}\right\}, \tag{11.118}$$

so k approaches k_B rapidly as $R \to \infty$ along lines where $\cos n\psi = 0$. Thus the contours of constant phase, Θ, are approximately curves of constant ψ. Since we also have from equations (11.116) and (11.117) that $\phi = \psi + \pi/2$ ($s_{n0} < 0$) or $\phi = \psi - \pi/2$ ($s_{n0} > 0$), the phase contours will be lines of constant ϕ. The pattern therefore consists of patches of straight lines radiating from the defect core along directions such that $\cos n\psi = 0$, joined together by regions of curved rolls. The special case $n = 2$, the saddle, is shown in Figure 11.4b, and the concave

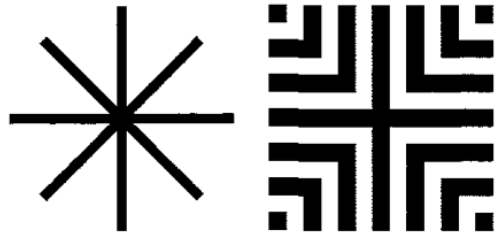


Fig 11 4. Sketches of the phase contours in the vicinity of (a) the vortex and (b) the saddle singularities of the stationary Cross–Newell equations

disclination, where $n = 3$, is shown in Figure 11 3b. Both sketches show idealised versions of the defects without curved joining regions. Note that for n odd only half the values of ψ such that $\cos n\psi = 0$ give rise to straight line patches. This is because the rapid approach to $k = k_B$ depends on the signs of X and Y in the function $k(X \sin\psi - Y\cos\psi)/s_{n0}n\sin n\psi$, which for n odd is unchanged under the transformation $\psi \to \psi + \pi$, while for n even it changes sign. Thus for n odd, the two directions ψ and $\psi + \pi$ give a rapid approach for the same signs of X and Y, thus defining only one straight line, whereas for n even the two directions pick out opposite signs of X and Y and hence define two lines.

Further details of all the point defect solutions can be found in Passot and Newell (1994) and Ercolani *et al* (2000)

11.4 Models with variational structure

If the original governing equation has variational structure such that

$$\frac{\partial u}{\partial t} = -\frac{\delta \mathcal{F}}{\delta u}, \tag{11.119}$$

where

$$\mathcal{F}(u) = \int_{\mathcal{D}} f(u)\mathrm{d}x\mathrm{d}y, \tag{11.120}$$

and where $\mathcal{D}$ is the domain of pattern formation and u and $f(u)$ are real, then the analysis takes a particularly simple form

To leading order the following equations hold

$$\frac{\partial u}{\partial t} = \epsilon \frac{\partial \Theta}{\partial T}\frac{\partial u_0}{\partial \theta}, \tag{11.121}$$

$$\delta u = \frac{\partial u_0}{\partial \theta}\delta\theta \tag{11.122}$$

If we now rearrange the governing equation (11.119) to give

$$\int_{\mathcal{D}} \frac{\partial u}{\partial t} \delta u \, \mathrm{d}x\mathrm{d}y = -\delta \mathcal{F}, \tag{11.123}$$

where the integral is needed to take account of the functional derivative in equation (11.119), and then substitute from equations (11.121) and (11.122), and average u_0 and its derivatives over θ we get

$$\epsilon \frac{\partial \Theta}{\partial T} \int_{\mathcal{D}} \overline{\left(\frac{\partial u_0}{\partial \theta} \right)^2} \delta\theta \, \mathrm{d}x\mathrm{d}y = -\delta \overline{\mathcal{F}}, \tag{11.124}$$

where $\overline{(\,)}$ denotes the averaging. Now since the system is translationally and rotationally symmetric, the averaged free energy density, $\bar{f}$, can only depend on k^2. Thus we also have

$$\delta \overline{\mathcal{F}} = \int_{\mathcal{D}} \frac{\mathrm{d}\bar{f}}{\mathrm{d}k^2} \delta k^2 \, \mathrm{d}x\mathrm{d}y, \tag{11.125}$$

and so

$$\epsilon \frac{\partial \Theta}{\partial T} \int_{\mathcal{D}} \overline{\left(\frac{\partial u_0}{\partial \theta} \right)^2} \delta\theta \, \mathrm{d}x\mathrm{d}y = -\int_{\mathcal{D}} \frac{\mathrm{d}\bar{f}}{\mathrm{d}k^2} \delta k^2 \, \mathrm{d}x\mathrm{d}y \tag{11.126}$$

Integrating over the domain, $\mathcal{D}$, and remarking that

$$\delta k^2 = \delta(\nabla_x \theta \quad \nabla_x \theta) = 2\nabla_x \theta \quad \nabla_x \delta\theta = 2\boldsymbol{k} \quad \nabla_x \delta\theta, \tag{11.127}$$

the divergence theorem can be used with suitable boundary conditions, to show that

$$\begin{aligned} &\epsilon \frac{\partial \Theta}{\partial T} \int_{\mathcal{D}} \overline{\left(\frac{\partial u_0}{\partial \theta} \right)^2} \delta\theta \, \mathrm{d}x\mathrm{d}y \\ &= \epsilon \int_{\mathcal{D}} \nabla \cdot \left\{ 2\boldsymbol{k} \frac{\mathrm{d}\bar{f}}{\mathrm{d}k^2} \right\} \delta\theta \, \mathrm{d}x\mathrm{d}y, \end{aligned} \tag{11.128}$$

where

$$\nabla \equiv \nabla_X = \epsilon^{-1} \left(\nabla_x - \boldsymbol{k} \frac{\partial}{\partial \theta} \right), \tag{11.129}$$

and we have used the fact that neither $\boldsymbol{k}$ nor $\bar{f}$ has any explicit (short-scale) θ dependence. Since $\delta\theta$ is arbitrary, we can extract the phase equation

$$\frac{\partial \Theta}{\partial T} \overline{\left(\frac{\partial u_0}{\partial \theta} \right)^2} = \nabla \cdot \left(2\boldsymbol{k} \frac{\mathrm{d}\bar{f}}{\mathrm{d}k^2} \right) \tag{11.130}$$

This equation can be written in the form

$$\tau(k)\frac{\partial \Theta}{\partial T} + \nabla \cdot (\boldsymbol{k}B(k)) = 0, \tag{11.131}$$

where we identify

$$\tau(k) = \overline{\left(\frac{\partial u_0}{\partial \theta}\right)^2} \tag{11.132}$$

$$B(k) = -2\frac{\mathrm{d}\overline{f}}{\mathrm{d}k^2} \tag{11.133}$$

Hence we have

$$\overline{\mathcal{F}} = \frac{1}{\epsilon^2}\int_{\mathcal{D}}\left(-\frac{1}{2}\int^{k^2}\widetilde{B}(z)\mathrm{d}z\right)\mathrm{d}X\mathrm{d}Y, \tag{11.134}$$

where $\widetilde{B}(k^2) \equiv B(k)$, which is just the same as the formal free energy defined in equation (11.29) apart from a rescaling by ϵ^{-2}. Recall, however, that in Section 11.1 we did not assume the existence of a free energy, $\mathcal{F}$, for the governing equation, and indeed there need not be one.

To see how this all works in practice we need to take a particular model equation and find the appropriate form of $f(u)$. Let's have a go at our usual example, the Swift–Hohenberg equation:

$$\frac{\partial u}{\partial t} = \mu u - u^3 - (\nabla^2 + 1)^2 u. \tag{11.135}$$

In this case the free energy density, $f(u)$, takes the form

$$f(u) = -\frac{1}{2}\mu u^2 + \frac{1}{4}u^4 + \frac{1}{2}[(\nabla^2 + 1)u]^2 \tag{11.136}$$

At leading order the equation for u_0 is given by

$$\mu u_0 - u_0^3 - \left(1 + k^2\frac{\partial^2}{\partial \theta^2}\right)^2 u_0 = 0, \tag{11.137}$$

and if we assume that u_0 takes the form

$$u_0 = A\cos\theta + h.o.t., \tag{11.138}$$

where A is a constant, then substituting into equation (11.137) we get

$$\begin{aligned}&\mu A\cos\theta - (1-k^2)^2 A\cos\theta - A^3\cos^3\theta = \\ &\mu A\cos\theta - (1-k^2)^2 A\cos\theta - A^3\left(\frac{3}{4}\cos\theta + \frac{1}{4}\cos 3\theta\right) = 0\end{aligned} \tag{11.139}$$

Neglecting the higher order term in $\cos 3\theta$ this tells us that

$$A^2 = \frac{4}{3}[\mu - (1-k^2)^2] \tag{11.140}$$

Now to leading order $\overline{f}$ is given by

$$\begin{aligned}
\overline{f} &= \frac{1}{2\pi}\int_0^{2\pi}\left(-\frac{1}{2}\mu u_0^2 + \frac{1}{4}u_0^4 + \frac{1}{2}\left[\left(1+k^2\frac{\partial^2}{\partial\theta^2}\right)u_0\right]^2\right)\mathrm{d}\theta,\\
&= \frac{1}{2\pi}\int_0^{2\pi}\left(-\frac{1}{2}\mu u_0^2 + \frac{1}{4}u_0^4 + \frac{1}{2}u_0\left[\left(1+k^2\frac{\partial^2}{\partial\theta^2}\right)\right]^2 u_0\right)\mathrm{d}\theta, \quad \text{(by parts)}\\
&= \frac{1}{2\pi}\int_0^{2\pi}\left(\frac{1}{4}u_0^4 - \frac{1}{2}u_0^4\right)\mathrm{d}\theta, \quad \text{(using the Swift–Hohenberg equation)}\\
&= -\frac{1}{8\pi}\int_0^{2\pi} u_0^4\,\mathrm{d}\theta,\\
&= -\frac{2}{9\pi}[\mu-(1-k^2)^2]^2\int_0^{2\pi}\cos^4\theta\,\mathrm{d}\theta,\\
&= -\frac{1}{6}[\mu-(1-k^2)^2]^2
\end{aligned} \tag{11.141}$$

From this we can find the form of $B(k)$:

$$B(k) = -2\frac{\mathrm{d}\overline{f}}{\mathrm{d}k^2} = \frac{4}{3}(\mu-(1-k^2)^2)(1-k^2) \tag{11.142}$$

We can also easily calculate $\tau(k)$, since

$$\tau(k) = \overline{\left(\frac{\partial u_0}{\partial\theta}\right)^2} = \frac{1}{2\pi}\int_0^{2\pi} A^2\sin^2\theta\,\mathrm{d}\theta = \frac{A^2}{2} = \frac{2}{3}[\mu-(1-k^2)^2], \tag{11.143}$$

and so it is possible to write down the phase equation, if desired. Instead we note that $\tau(k)$ is positive when $A \neq 0$, and use the form of $B(k)$ to determine the phase instability boundaries directly.

If we set $B(k) = 0$ we find that either $\mu = (1-k^2)^2$, which gives the neutral stability curve, or $k^2 = 1$, which is the zigzag instability boundary. The pattern is zigzag-unstable when $B(k) > 0$. Since we must be in the region above the neutral stability curve, we must have $\mu > (1-k^2)^2$ and the zigzag instability will set in if $k^2 < 1$. The critical wavenumber is $k = 1$; if we allow small departures from criticality by setting $k = 1+q$, with $|q| \ll 1$ we can retrieve these boundaries in

more familiar form: the neutral stability curve becomes $\mu = 4q^2$ (note the rescaling of q relative to Chapter 7) and the zigzag instability is found to occur when $q < 0$. Strictly speaking the Cross–Newell analysis is not valid close to the neutral stability boundary, as then the pattern amplitude is small and so we are not in the fully nonlinear regime. None the less it is encouraging that the results agree with those found in Chapter 8 for patterns close to onset.

The pattern is Eckhaus-unstable when $\mathrm{d}(kB)/\mathrm{d}k > 0$. It is straightforward to work out that this criterion takes the form

$$\frac{\mathrm{d}}{\mathrm{d}k}(kB) = \frac{4}{3}[\mu(1-3k^2) - (1-k^2)^2(1-7k^2)] > 0 \tag{11.144}$$

Substituting $k = 1 + q$ with $|q| \ll 1$ we find that the Eckhaus instability sets in for $\mu < 12q^2$, which once again agrees reassuringly with the results of Chapter 8, taking into account the rescaling of q.

11.5 Systems with mean drift

There are some systems for which the Cross–Newell reduction does not result in a smooth gradient expansion in the phase Θ, because variations in the wavevector lead to a mean drift, which in turn advects the phase. The most notable case for which this is true is that of Boussinesq Rayleigh–Bénard convection at low to moderate Prandtl numbers (see Section 8.6). Cross and Newell (1984) considered the effect of mean drift using a phenomenological extension of the Cross–Newell equation:

$$\frac{\partial \Theta}{\partial T} = -\tau^{-1}(k)\nabla \quad (\boldsymbol{k}B(k)) - \boldsymbol{U} \cdot \boldsymbol{k}, \tag{11.145}$$

where

$$\boldsymbol{U} = \left(\frac{\partial \psi}{\partial Y}, -\frac{\partial \psi}{\partial X}\right), \tag{11.146}$$

and

$$\nabla^2 \psi = \gamma \widehat{\boldsymbol{z}} \quad \nabla \times (\boldsymbol{k}\nabla \cdot (\boldsymbol{k}A^2)), \tag{11.147}$$

where $\widehat{\boldsymbol{z}}$ is the unit vector normal to the horizontal plane of pattern formation, γ is a coupling constant (the inverse Prandtl number in convection) and $A \equiv A(k^2)$ is the pattern amplitude. Note that $\boldsymbol{U}$ is solenoidal ($\nabla \cdot \boldsymbol{U} = 0$) and depends nonlocally on the phase via the variable ψ since we must integrate equation (11.147) over the two horizontal directions to find ψ. This modified form of the phase equation was inspired by an analysis of the Boussinesq convection equations. The

flow field $\boldsymbol{U}$ represents the singular component of the drift, while any nonsingular part is included in the term involving $B(k)$. Deriving the equations directly from the full Oberbeck–Boussinesq convection equations, Newell, Passot and Souli (1990) showed that in fact the velocity $\boldsymbol{U}$ in equation (11.145) should be multiplied by a weighting function $\rho(k)$ that accounts for the averaging of the mean drift over the depth of the convection layer, and also that the equation for the streamfunction, ψ, is much more complicated than equation (11.147). However, although there are quantitative differences, the main ideas are captured by the simpler theory, so we shall describe these briefly.

If we consider perturbations to a straight roll pattern, so that $\boldsymbol{k} = (k, 0) + \nabla\widetilde{\Theta}$, where $|\widetilde{\Theta}| \ll 1$ represents the perturbations, then at leading order in $\widetilde{\Theta}$ we find that ψ and $\widetilde{\Theta}$ are linearly related according to

$$\nabla^2\psi = \gamma\left(k^2\frac{\mathrm{d}A^2}{\mathrm{d}k}\frac{\partial^3\widetilde{\Theta}}{\partial X^2\partial Y} + kA^2\nabla^2\left(\frac{\partial\widetilde{\Theta}}{\partial Y}\right)\right), \tag{11.148}$$

and at leading order in $\widetilde{\Theta}$ the phase equation is given by

$$\frac{\partial\widetilde{\Theta}}{\partial T} = D_\parallel(k)\frac{\partial^2\widetilde{\Theta}}{\partial X^2} + D_\perp(k)\frac{\partial^2\widetilde{\Theta}}{\partial Y^2} - k\frac{\partial\psi}{\partial Y}, \tag{11.149}$$

where the diffusion coefficients are as given in equations (11.21) and (11.22).

Now if $\psi = \widehat{\psi}e^{\sigma T + iq\cos\phi X + iq\sin\phi Y}$ and $\widetilde{\Theta} = \widehat{\Theta}e^{\sigma T + iq\cos\phi X + iq\sin\phi Y}$, where $\widehat{\psi}$ and $\widehat{\Theta}$ are constants, so that we are looking at disturbances with wavevector $\boldsymbol{q}$ at an angle ϕ to the original wavevector $\boldsymbol{k}$, then we find from equations (11.148) and (11.149) that the growth rate σ of such perturbations is given by

$$\sigma = -q^2(D_\parallel^{\mathrm{eff}}\cos^2\phi + D_\perp^{\mathrm{eff}}\sin^2\phi), \tag{11.150}$$

where

$$D_\parallel^{\mathrm{eff}} = D_\parallel + k^2\gamma\frac{\mathrm{d}(kA^2)}{\mathrm{d}k}\sin^2\phi, \tag{11.151}$$

$$D_\perp^{\mathrm{eff}} = D_\perp + k^2\gamma A^2\sin^2\phi. \tag{11.152}$$

The zigzag instability to modes at rightangles to the original wavevector is given by $\sin\phi = 1$ and sets in at $D_\perp^{\mathrm{eff}} = 0$ or $D_\perp + k^2\gamma A^2 = 0$ which is equivalent to $B(k) - \gamma\tau k^2A^2 = 0$. The zigzag instability is therefore stabilised by the presence of mean drift if, as expected, γ is positive, since then the effective diffusion coefficient is greater than $D_\perp$.

The Eckhaus instability to modes parallel to the original wavevector occurs for $\sin\phi = 0$ and sets in at $D_{\|}^{\text{eff}} = D_{\|} = 0$, the boundary being unaffected by the mean drift.

At intermediate angles ϕ we find the skew-varicose instability, which arises from destabilisation of the Eckhaus mode by the mean drift: in the case $\mathrm{d}(kA^2)/\mathrm{d}k < 0$, the effective diffusion coefficient decreases below $D_{\|}$. The skew-varicose instability can be thought of then as an Eckhaus instability with transverse modulations brought about by the mean drift. There are some changes to the position of the skew-varicose instability boundary in the more accurate theory introduced by Newell, Passot and Souli (1990).

It would be helpful to know whether mean drift affects wavenumber selection. In the vicinity of an axisymmetric focus singularity, the pattern has radial symmetry, and so $\boldsymbol{U} \equiv U(r)\hat{\boldsymbol{r}}$, where $\hat{\boldsymbol{r}}$ is a unit vector in the radial direction. Since $\nabla \cdot \boldsymbol{U} = 0$, this implies that $U(r) \propto 1/r$ which is not possible in the absence of sources or sinks of fluid at $r = 0$. So $\boldsymbol{U}$ must be zero and the wavenumber selection effect that was found in the absence of mean drift still holds. The selected wavenumber k_0 still satisfies $B(k_0) = 0$, but k_0 is no longer equal to the wavenumber k_z at which the zigzag instability sets in, which now satisfies

$$B(k_z) - \gamma\tau(k_z)k_z^2 A^2(k_z) = 0. \tag{11.153}$$

For nonaxisymmetric foci, however, $\boldsymbol{U}$ is nonzero and since $\boldsymbol{U}$ is nonlocal in the phase gradients the wavevector is no longer determined independently along each trajectory orthogonal to the roll axes. Cross and Newell (1984) suggested that integrating over a closed contour defined by a closed roll loop, there can be no net contribution from $\boldsymbol{U}$ since the flow is incompressible ($\nabla \cdot \boldsymbol{U} = 0$), and that therefore we can expect the wavenumber averaged around a roll contour to be roughly k_0 at large distances from the focus. When the weighting $\rho(k)$ is included, the depth-averaged mean drift, $\rho(k)\boldsymbol{U}$, that advects the phase contours, is not solenoidal ($\nabla \cdot (\rho(k)\boldsymbol{U}) \neq 0$), although Newell, Passot and Souli (1990) found $\rho(k)$ to be almost constant. In any case the picture here is less clear. Newell, Passot and Souli (1990) went on to look at the destabilisation of target patterns by the focus instability, which they predicted to take place well within the Busse stability balloon and therefore to be important in the onset of time-dependent, spatially disordered states, such as spiral defect chaos, described in Chapter 10.

This is the end of the book. If it has piqued your curiosity and you want to find out more about patterns, there is an extensive research literature to explore – the references provide a starting point.

Exercises

11 1 Work out the forms of $\tau(k)$ and $B(k)$ for the variational governing equation

$$\frac{\partial u}{\partial t} = \mu u - u^3 + a(u|\nabla u|^2 + u^2\nabla^2 u) - (\nabla^2 + 1)^2 u,$$

where μ is a real bifurcation parameter and a is a real constant. Hence find an equation for the zigzag instability boundary in terms of μ and k^2.

11 2 If $u_0 = A(k)\cos\theta$, find $A(k)$ when u is governed by the nonvariational equation

$$\frac{\partial u}{\partial t} = \mu u - u^3 + au|\nabla u|^2 + bu^2\nabla^2 u - (\nabla^2 + 1)^2 u,$$

where μ is a real bifurcation parameter and a and b are real constants with $a \neq b$. Now find the equation for $\partial\Theta/\partial T$ in terms of $A(k)$ and $\boldsymbol{k}$, using equation (11 12)

References

Acheson, D J. (1990) *Elementary Fluid Dynamics* (Oxford: Clarendon)

Agladze, K. I , Panfilov, A V and Rudenko, A N (1988) Nonstationary rotation of spiral waves: three-dimensional effect, *Physica D.* **29**, 409–15

Arbell, H and Fineberg, J. (2002) Pattern formation in two-frequency forced parametric waves, *Phys Rev. E* **65**, 036224

Arfken, G. B (2001) *Mathematical Methods for Physicists*, 5th edition (London: Academic).

Arnol'd, V I. (1983) *Geometrical Methods in the Theory of Ordinary Differential Equations* (New York: Springer)

Aronson, D G and Weinberger, H F (1978) Multidimensional nonlinear diffusion arising in population genetics, *Adv. Math.* **30**, 33–76

Bär, M , Hildebrand, M., Eiswirth, M , Falcke, M., Engel, H. and Neufeld, M (1994) Chemical turbulence and standing waves in a surface reaction model: the influence of global coupling and wave instabilities, *Chaos* **4**, 499–508

Barkley, D. (1994) Euclidean symmetry and the dynamics of rotating spiral waves, *Phys Rev. Lett.* **72**, 164–7.

(1995) Spiral meandering, in *Chemical Waves and Patterns*, R Kapral and K Showalter eds (Dordrecht: Kluwer).

Belmonte, A., Flesselles, J -M and Ouyang, Q. (1996) Spiral instability to line sources in forced chemical pattern turbulence, *Europhys Lett.* **35**, 665–70.

Belousov, B P. (1958) A periodic reaction and its mechanism, *Sbornik Referatov po Radiatsionnoĭ Meditsine za 1958 god* (Moscow: Medgiz), p. 145

Bénard, H. (1900) Les tourbillons cellulaires dans une nappe liquide Première partie: description générale des phénomènes, *Rev. Gen Sci Pures Ap* **11**, 1261–71

Benjamin, T B and Feir, J. E (1967) The disintegration of wave trains on deep water, *J Fluid Mech.* **27**, 417–30

Bernoff, A J (1994) Finite-amplitude convection between stress-free boundaries: Ginzburg-Landau equations and modulation theory, *Euro Jnl of Applied Mathematics* **5**, 267–82.

Bestehorn, M (1993) Phase and amplitude instabilities for Bénard–Marangoni convection in fluid layers with large aspect ratio, *Phys Rev. E* **48**, 3622–34.

(1994) Pattern selection in Bénard–Marangoni convection, *Int J Bif. Chaos* **4**, 1085–94.

Billingham, J. and King, A. C. (2000) *Wave Motion* (Cambridge: Cambridge University Press).

Bodenschatz, E., Pesch, W and Ahlers, G. (2000) Recent developments in Rayleigh–Bénard convection, *Annu. Rev. Fluid Mech* , **32**, 709–78

Bodenschatz, E , Pesch, W and Kramer, L (1988) Structure and dynamics of dislocations in anisotropic pattern-forming systems, *Physica D.* **32**, 135–45

Bosch Vivancos, I , Chossat, P. and Melbourne, I (1995) New planforms in systems of partial differential equations with Euclidean symmetry, *Arch Rational Mech. Anal* **131**, 199–224.

Bressloff, P. C , Cowan, J. D , Golubitsky, M and Thomas, P. J. (2001a) Scalar and pseudoscalar bifurcations motivated by pattern formation on the visual cortex, *Nonlinearity* **14**, 739–75

Bressloff, P. C , Cowan, J. D., Golubitsky, M , Thomas, P. J and Wiener, M C. (2001b) Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex, *Phil. Trans. R. Soc Lond* **B 356**, 299–330.

Bröcker, T and tom Dieck, T (1985) *Representations of Compact Lie Groups* (New York: Springer-Verlag)

Busse, F H (1962) *Das Stabilitätsverhalten der Zellularkonvektion bei endlicher Amplitude*, Ph.D thesis (München: Ludwig-Maximilians-Universität).

(1967) The stability of finite amplitude cellular convection and its relation to an extremum principle, *J Fluid Mech.* **30**, 625–9.

(1972) The oscillatory instability of convection rolls in a low Prandtl number fluid, *J Fluid Mech* **52**, 97–112.

Busse, F. H and Bolton, E. W. (1984) Instabilities of convection rolls with stress-free boundaries near threshold, *J Fluid Mech* **146**, 115–25

Busse, F. H and Clever, R M (1979) Instabilities of convection rolls in a fluid of moderate Prandtl number, *J. Fluid Mech* **91**, 319–35

Busse, F H. and Heikes, K. E. (1980) Convection in a rotating layer: a simple case of turbulence, *Science* **208**, 173–5

Busse, F H. and Whitehead, J. A. (1971) Instabilities of convection rolls in a high Prandtl number fluid, *J. Fluid Mech.*, **47**, 305–20

Buzano, E and Golubitsky, M. (1983) Bifurcation involving the hexagonal lattice and the planar Bénard problem, *Phil Trans. Roy. Soc Lond A* **308**, 617–67

Cakmur, R V, Egolf, D A , Plapp, B B and Bodenschatz, E (1997) Bistability and competition of spatiotemporal chaotic and fixed point attractors in Rayleigh–Bénard convection, *Phys Rev. Lett* **79**, 1853–6.

Carr, J (1981), *Applications of Centre Manifold Theory* (New York: Springer-Verlag)

Castets, V , Dulos, E , Boissonade, J and De Kepper, P. (1990) Experimental evidence of a sustained standing Turing-type nonequilibrium chemical pattern, *Phys. Rev. Lett* **64**, 2953–6

Chiam, K -H , Paul, M. R , Cross, M C and Greenside, H. S (2003) Mean flow and spiral defect chaos in Rayleigh–Bénard convection, *Phys. Rev. E* **67**, 056206

Chossat, P. and Iooss, G (1994) *The Couette–Taylor Problem* (New York: Springer-Verlag)

Chossat, P. and Lauterbach, R. (2000) *Methods in Equivariant Bifurcations and Dynamical Systems* (Singapore: World Scientific)

Cicogna, G. (1981) Symmetry breakdown from bifurcations, *Lettere al Nuovo Cimento* **31**, 600–2.

Ciliberto, S., Coullet, P., Lega, J., Pampaloni, E and Peres-Garcia, C (1990) Defects in roll-hexagon competition, *Phys Rev. Lett.* **65**, 2370–3

Cohn, P. M. (1982) *Algebra*, Volume 1, 2nd edn. (Chichester: Wiley)

Cornwell, J. F. (1984) *Group Theory in Physics*, Volume I (London: Academic)

Coullet, P. and Fauve, S. (1985) Propagative phase dynamics for systems with Galilean invariance *Phys. Rev. Lett.* **55**, 2857–9

Coullet, P., Gil, L. and Lega, J. (1989) Defect-mediated turbulence, *Phys. Rev. Lett.* **62**, 1619–22

Crawford, J. D. (1994) $D_4 + T^2$ mode interactions and hidden rotational symmetry, *Nonlinearity* **7**, 697–739

Crawford, J. D., Golubitsky, M., Gomes, M. G. M., Knobloch, E. and Stewart, I. (1991) Boundary conditions as symmetry constraints, in *Singularity Theory and Its Applications*, Warwick 1989 (R. M. Roberts and I. N. Stewart) Lecture Notes in Mathematics **2** (Heidelberg: Springer-Verlag)

Cross, M. C. (1982) Ingredients of a theory of convective textures close to onset, *Phys. Rev. A* **25**, 1065–76

Cross, M. C. and Hohenberg, P. C. (1993) Pattern formation outside of equilibrium, *Rev. Mod. Phys.* **65**, 851–1112

Cross, M. C. and Newell, A. C. (1984) Convection patterns in large aspect ratio systems, *Physica D* **10**, 299–328

Daniels, K. E. and Bodenschatz, E. (2003) Statistics of defect motion in spatiotemporal chaos in inclined layer convection, *Chaos* **13**, 55–63

Decker, W. and Pesch, W. (1994) Order parameter and amplitude equations for the Rayleigh–Bénard convection, *J. Phys. II France* **4**, 419–38

Dellnitz, M., Field, M., Golubitsky, M., Hohmann, A. and Jun Ma (1995) 'Cycling chaos' IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, **42**, 821–3

Dionne, B. and Golubitsky, M. (1992) Planforms in two and three dimensions, *Z. Angew. Math. Phys.* **43**, 36–62.

Dionne, B., Silber, M. and Skeldon, A. C. (1997) Stability results for steady, spatially periodic planforms, *Nonlinearity* **10**, 321–53.

Doelman, A., Sandstede, B., Scheel, A. and Schneider, G. (2003) Propagation of hexagonal patterns near onset, *Euro. Jnl of App. Math.* **14**, 85–110

Dubois, M., Bergé, P. and Wesfreid, J. (1978) Non Boussinesq convective structures in water near 4° C, *J. Phys. (Paris)* **39**, 1253–7

Echebarria, B. and Pérez-García, C. (1998) Phase instabilities in hexagonal patterns, *Europhys. Lett.* **43**, 35–40.

(2001) Stability of hexagonal patterns in Bénard–Marangoni convection, *Phys. Rev. E* **63**, 066307

Echebarria, B. and Riecke, H. (2001) Sideband instabilities and defects of quasipatterns, *Physica D.* **158**, 45–68.

Eckhaus, W. (1965) *Studies in Nonlinear Stability Theory* (Berlin: Springer-Verlag)

Edwards, W. S. and Fauve, S. (1993) Parametrically excited quasi-crystalline surface-waves, *Phys. Rev. E* **47**, R788–91.

Egolf, D. A., Melnikov, I. V. and Bodenschatz, E. (1998) Importance of local pattern properties in spiral defect chaos, *Phys. Rev. Lett.* **80**, 3228–31.

Ercolani, N. M., Indik, R., Newell, A. C. and Passot, T. (2000) The geometry of the phase diffusion equation, *J. Nonlinear Sci.* **10**, 223–74

(2003) Global description of patterns far from onset: a case study, *Physica D* **184**, 127–40

Faraday, M. (1831) On a peculiar class of acoustical figures; and on certain forms assumed by groups of particles upon vibrating elastic surfaces, *Phil. Trans. R. Soc. Lond.* **121**, 299–340.

Fauve, S (1991) 'Woods Hole GFD summer school notes: principal lectures', Woods Hole Oceanographical Institute Technical Report WHOI-92-16

(1998) Pattern forming instabilities *in Hydrodynamics and Nonlinear Instabilities* (eds C Godrèche and P. Manneville), (Cambridge: Cambridge University Press).

Fauve, S , Bolton, E. W. and Brachet, M. E (1987) Nonlinear oscillatory convection: a quantitative phase dynamics approach, *Physica D* **29**, 202–14.

Fauve, S., Douady, S. and Thual, O (1991) Drift instabilities of cellular patterns, *J Phys II (Paris)* **1**, 311–22

FitzHugh, R. (1961) Impulses and physiological states in theoretical models of nerve membrane, *Biophys. J* **1**, 445–66.

Ginzburg, V. L and Landau, L. D. (1950) On the theory of superconductivity, *Zh Eksperim. I. Teor. Fiz.* **20**, 1064–82.

Glendinning, P. (1994) *Stability, Instability and Chaos An Introduction to the Theory of Nonlinear Differential Equations* (Cambridge: Cambridge University Press)

Goldstein, H. F , Knobloch, E. and Silber, M. (1990) Planform selection in rotating convection, *Physics of Fluids A* **2**, 625–7

Golubitsky, M., Marsden J. and Schaeffer, D. (1984) Bifurcation problems with hidden symmetries, in *Partial Differential Equations and Dynamical Systems* (W E. Fitzgibbon, ed), *Research Notes in Mathematics* **101** (San Francisco: Pitman), 181–210.

Golubitsky, M. and Stewart, I (2002) *The Symmetry Perspective From Equilibrium to Chaos in Phase Space and Physical Space* (Basel: Birkhauser).

Golubitsky, M., Stewart, I and Schaeffer, D. G. (1988) *Singularities and Groups in Bifurcation Theory*, Volume II, (New York: Springer-Verlag).

Golubitsky, M., Swift, J. W and Knobloch, E (1984) Symmetries and pattern selection in Rayleigh–Bénard convection, *Physica D* **10**, 249–76

Gomes, M G M , Labouriau, I S. and Pinho, E. M (1999) Spatial hidden symmetries in pattern formation, in *Pattern Formation in Continuous and Coupled Systems* (M. Golubitsky, D Luss and S. Strogatz, eds.) Springer-Verlag IMA Volumes in Mathematics and its Applications **115**, 83–100.

Guckenheimer, J and Holmes, P. (1983) *Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Fields* (New York: Springer-Verlag).

(1988) Structurally stable heteroclinic cycles, *Math Proc Camb Phil. Soc.* **103**, 189–92

Hagan, P. S. (1982) Spiral waves in reaction-diffusion equations, *SIAM J. Appl. Math* **42**, 762–86

Hagan, P. S. and Cohen, M. S (1981) Diffusion-induced morphogenesis in the development of *Dictyostelium*, *J Theor. Biol.* **93**, 881–908

Haken, H (1978) *Synergetics – An Introduction* (Berlin: Springer-Verlag)

Hendry, M , Nam, K , Guzdar, P. and Ott, E (2000) Target waves in the complex Ginzburg–Landau equation, *Phys. Rev. E* **62**, 7627–31

Hirsch, M. W and Smale, S (1974) *Differential Equations, Dynamical Systems and Linear Algebra* (New York and London: Academic).

Hoyle, R. B. (1993) Long wavelength instabilities of square patterns, *Physica D* **67**, 198–223

(1994a) *Instabilities of Three-dimensional Patterns*, Ph D. thesis (Cambridge: University of Cambridge)

(1994b) Phase instabilities of oscillatory standing squares and alternating rolls, *Phys. Rev. E* **49**, 2875–80.

(1995) Nonlinear phase diffusion equations for the longwave instabilities of hexagons, *Appl Math Lett.* **8**:3, 81–5
(1998a) Zigzag and Eckhaus instabilities in a quintic-order nonvariational Ginzburg–Landau equation, *Phys. Rev. E* **58**, 7315–18
(1998b) Universal instabilities of rolls, squares and hexagons, in *Time-dependent nonlinear convection*, *Adv. Fluid Mech.* **19**, 51–82. (Southampton: Computer Mechanics).
Jahnke, W., Skaggs, W. E. and Winfree, A. T. (1989) Chemical vortex dynamics in the Belousov–Zhabotinsky reaction and in the two-variable Oregonator model, *J. Phys Chem* **93**, 740–9.
James, G. and Liebeck, M. (1993) *Representations and Characters of Groups* (Cambridge: Cambridge University Press)
Johnston, B. L. and Richman, F. (1997) *Numbers and Symmetry* (Boca Raton: CRC Press).
Kawasaki, K. (1984) Defect-phase dynamics for dissipative media with potential, *Prog. Theor. Phys. Suppl.* **80**, 123–38
Kim, D. T., Kwan, Y., Lee, J. J., Ikeda, T., Uchida, T., Kamjoo, K., Kim, Y.-H., Ong, J. J. C., Athill, C. A., Wu, T.-J., Czer, L., Karagueuzian, H. S. and Chen, P.-S. (1998) Patterns of spiral tip motion in cardiac tissues, *Chaos* **8**, 137–48
Kirk, V. and Silber, M. (1994) A competition between heteroclinic cycles, *Nonlinearity* **7**, 1605–21.
Komarova, N. L. and Newell, A. C. (2000) Nonlinear dynamics of sand banks and sand waves, *J. Fluid Mech.* **415**, 285–321
Kramer, L., Ben-Jacob, E., Brand, H. and Cross M. C. (1982) Wavelength selection in systems far from equilibrium, *Phys. Rev. Lett.* **49**, 1891–4.
Kramer, L. and Zimmermann W. (1985) On the Eckhaus instability for spatially periodic patterns, *Physica D* **16**, 221–32.
Krishnamurti, R. (1970) On the transition to turbulent convection. Part 2: The transition to time-dependent flow, *J. Fluid Mech.* **42**, 309–20.
(1973) Some further studies on the transition to turbulent convection, *J. Fluid Mech.* **60**, 285–303.
Kudrolli, A., Pier, B. and Gollub, J. P. (1998) Superlattice patterns in surface waves, *Physica D* **123**, 99–111.
Kuecken, M. U. (2004) *On the Formation of Fingerprints*, Ph. D. thesis (Tucson: University of Arizona)
Kumar, K. (1996) Linear theory of Faraday instability in viscous liquids, *Proc. R. Soc. Lond. A* **452**, 1113–26.
Küppers, G. and Lortz, D. (1969) Transition from laminar convection to thermal turbulence in a rotating fluid layer, *J. Fluid Mech.* **35**, 609–20.
Kuramoto, Y. (1984a) Phase dynamics of weakly unstable periodic structures, *Prog. Theor. Phys.* **71**, 1182–96
(1984b) *Chemical Oscillations, Waves and Turbulence* (New York: Springer-Verlag)
Kuramoto, Y. and Tsuzuki, T. (1976) Persistent propagation of concentration waves in dissipative media far from thermal equilibrium, *Prog. Theor. Phys.* **55**, 356–69
Kuske, R. and Milewski, P. (1999) Modulated two-dimensional patterns in reaction-diffusion systems, *Euro. Jnl of App. Math.* **10**, 157–84.
Kuznetsov, E. A., Nepomnyashchy, A. A. and Pismen, L. M. (1995) New amplitude equation for Boussinesq convection and nonequilateral hexagonal patterns, *Phys. Lett. A* **205**, 261–5.
Lauzeral, J., Metens, S. and Walgraef, D. (1993) On the phase dynamics of hexagonal patterns, *Europhys. Lett.* **24**, 707–12

Lee, K. J. (1997) Wave pattern selection in an excitable system, *Phys. Rev. Lett* **79**, 2907–10

Lee, K. J., Cox, E. C. and Goldstein, R. E. (1996) Competing patterns of signaling activity in *Dictyostelium discoideum*, *Phys. Rev. Lett* **76**, 1174–7

Li, G., Ouyang, Q., Petrov, V. and Swinney, H. L. (1996) Transition from simple rotating chemical spirals to meandering and traveling spirals, *Phys. Rev. Lett* **77**, 2105–8.

Malomed, B. A., Nepomnyaschchiĭ, A. A. and Tribelskiĭ, M. I. (1989) Two-dimensional quasiperiodic structures in nonequilibrium systems, *Sov. Phys. JETP* **69**, 388–96.

Malomed, B. A., Nepomnyashchy, A. A. and Tribelsky, M. I. (1990) Domain boundaries in convection patterns, *Phys. Rev. A* **42**, 7244–63.

Manneville, P. (1990) *Dissipative structures and weak turbulence* (San Diego: Academic)

Manoel, M. and Stewart, I. (2000) The classification of bifurcations with hidden symmetries, *P. Lond. Math. Soc* **80**, 198–234.

Matthews, P. C. (2003) Pattern formation on a sphere, *Phys. Rev. E* **67**, 036206

(2004) Automating symmetry-breaking calculations, *LMS J. Comp. & Math* **7**, 101–19

Matthews, P. C. and Cox, S. M. (2000a) One-dimensional pattern formation with Galilean invariance near a stationary bifurcation *Phys. Rev. E* **62**, R1473–6.

(2000b) Pattern formation with a conservation law *Nonlinearity* **13**, 1293–320

May, R. M. and Leonard, W. J. (1975) Nonlinear aspects of competition between three species, *SIAM J. Appl. Math* **29**, 243–53

Melbourne, I. (1998) Derivation of the time-dependent Ginzburg–Landau equation on the line, *J. Nonlinear Sci.* **8**, 1–15

(1999) Steady-state bifurcation with Euclidean symmetry, *T. Am. Math. Soc.* **351**, 1575–603

Melbourne, I., Chossat, P. and Golubitsky, M. (1989) Heteroclinic cycles involving periodic solutions in mode interactions with O(2) symmetry, *Proc. Roy. Soc. Edin.* **113A**, 315–45

Melnikov, I., V., Egolf, D. A., Jeanjean, S., Plapp, B. B. and Bodenschatz, E. (2000) Invasion of spiral defect chaos into straight rolls in Rayligh–Bénard convection, in *Stochastic Dynamics and Pattern Formation in Biological and Complex Systems* (eds S. Kim, K. J. Lee, T. K. Lim and W. Sung), AIP Conference Proceedings **501**(1), pp. 36–42

Melo, F., Umbanhowar, P. B. and Swinney, H. L. (1995) Hexagons, kinks and disorder in oscillated granular layers, *Phys. Rev. Lett* **75**, 3838–41.

Mielke, A. (1997) Mathematical analysis of sideband instabilities with application to Rayleigh–Bénard convection, *J. Nonlinear Sci.* **7**, 57–99

Morris, S. W., Bodenschatz, Cannell, D. S. and Ahlers, G. (1993) Spiral defect chaos in large aspect ration Rayleigh–Bénard convection, *Phys. Rev. Lett* **71**, 2026–9

Mullin, T. (2002) Mixing and de-mixing, *Science* **295**, 1851

Nagumo, J., Arimoto, S. and Yoshizawa, S. (1962) An active pulse transmission line simulating nerve axon, *Proc. IRE* **50**, 2061–70.

Nagy-Ungvarai, Z., Ungvarai, J. and Müller, S. C. (1993) Complexity in spiral wave dynamics, *Chaos* **3**, 15–19

Nettesheim, S., von Oertzen, A., Rotermund, H. H. and Ertl, G. (1993) Reaction diffusion patterns in the catalytic CO-oxidation on Pt(110): front propagation and spiral waves, *J. Chem. Phys.* **98**, 9977–85

Newell, A. C., Passot, T., Bowman, C., Ercolani, N. and Indik, R. (1996) Defects are weak and self-dual solutions of the Cross–Newell phase diffusion equation for natural patterns, *Physica D.* **97**, 185–205

Newell, A C , Passot, T and Souli, M. (1990) The phase diffusion and mean drift equations for convection at finite Rayleigh numbers in large containers, *J. Fluid Mech* **220**, 187–252

Newell, A C and Whitehead, J. A (1969) Finite bandwidth, finite amplitude convection, *J. Fluid Mech.* **38**, 279–303.

Norbury, J , Wei, J. and Winter, M. (2002) Existence and stability of singular patterns in a Ginzburg–Landau equation coupled with a mean field, *Nonlinearity* **15**, 2077–96

Ondarçuhu, T., Mindlin, G. B., Mancini, H. L. and Pérez García, C. (1993) Dynamical patterns in Bénard–Marangoni convection in a square container, *Phys Rev. Lett* **70**, 3892–5

Ouyang, Q and Flesselles, J -M (1996) Transition from spirals to defect turbulence driven by a convective instability, *Nature* **379**, 143–6.

Ouyang, Q , Swinney, H L. and Li, G (2000) Transition from spirals to defect-mediated turbulence driven by a Doppler instability, *Phys Rev. Lett* **84**, 1047–50

Panfilov, A V (1998) Spiral breakup as a model of ventricular fibrillation, *Chaos* **8**, 57–64

Passot, T and Newell, A. C. (1994) Towards a universal theory for natural patterns, *Physica D.* **74**, 301–52

Pearson, J. R. A. (1958) On convection cells induced by surface tension, *J. Fluid Mech* **4**, 489–500

Perez-Muñuzuri, V , Aliev, R , Vasiev, B., Perez-Villar, V and Krinsky, V I (1991) Super-spiral structures in an excitable medium, *Nature* **353**, 740–2.

Pismen, L. M (1981) Bifurcation into wave patterns and turbulence in reaction-diffusion equations, *Phys. Rev. A* **23**, 334–44

Plapp, B B., Egolf, D. A , Bodenschatz, E and Pesch, W (1998) Dynamics and selection of giant spirals in Rayleigh–Bénard convection, *Phys Rev. Lett* **81**, 5334–7

Plesser, T., Müller, S. C. and Hess, B. (1990) Spiral wave dynamics as a function of proton concentration in the ferroin-catalyzed Belousov–Zhabotinskii reaction, *J Phys. Chem* **94**, 7501–7.

Pocheau, A and Croquette, V (1984) Dislocation motion: a wavenumber selection mechanism in Rayleigh–Bénard convection, *J. Physique* **45**, 35–48.

Pollicott, S L , Matthews, P. C. and Cox, S. M (2003) Instability of convection in a fluid layer rotating about an oblique axis, *Phys Rev E* **67**, 016301.

Pomeau, Y and Manneville, P. (1979) Stability and fluctuations of a spatially periodic flow, *J Physique Lett* **40**, L609–12.

Pomeau, Y and Zaleski, S. (1983) Pattern selection in a slowly varying environment *J Physique Lett* **44**, L135–41

Pomeau, Y , Zaleski, S. and Manneville, P (1983) Dislocation motion in cellular structures *Phys Rev. A* **27**, 2710–26

Ponty, Y , Passot, T. and Sulem, P. L . (1997) Pattern dynamics in rotating convection at finite Prandtl number, *Phys Rev. E* **56**, 4162–78.

Porter, D. and Stirling, D S G (1990) *Integral Equations: A Practical Treatment from Spectral Theory to Applications* (Cambridge: Cambridge University Press)

Proctor, M R E (2001) Finite amplitude behaviour of the Matthews–Cox instability, *Phys. Lett. A* **292**, 181–7

Proctor, M R.E and Jones, C. A. (1988) The interaction of two spatially resonant patterns in thermal convection Part 1. Exact 1:2 resonance, *J Fluid Mech* **188**, 301–35

Rayleigh, Lord (1916) On convection currents in a horizontal layer of fluid, when the higher temperature is on the under side, *Philos Mag* **32**, 529–46.

Rucklidge, A. M. (2003) Pattern formation in large domains, *Phil Trans R Soc. Lond. A* **361**, 2649–64.
van Saarloos, W. (2003) Front propagation into unstable states, *Phys. Rep.* **386**, 29–222.
Sandstede, B. and Scheel, A (2000) Absolute versus convective instability of spiral waves, *Phys Rev. E* **62**, 7708–14.
(2001) Superspiral structures of meandering and drifting spiral waves, *Phys. Rev. Lett* **86**, 171–4
(2004) Defects in oscillatory media: towards a classification. *SIAM J. Applied Dynamical Systems* **3**, 1–68.
Sandstede, B., Scheel, A. and Wulff, C. (1997) Center-manifold reduction for spiral waves, *C R Acad Sci., Serie I, Math.* **324**, 153–8.
(1999a) Dynamical behavior of patterns with Euclidean symmetry, in *Pattern Formation in Continuous and Coupled Systems* (M Golubitsky, D. Luss and S. Strogatz, eds) Springer-Verlag, IMA Volumes in Mathematics and its Applications **115**, 249–64
(1999b) Bifurcations and dynamics of spiral waves. *J. Nonlinear Sci* **9**, 439–78.
Schlüter, A , Lortz, D and Busse, F H. (1965) On the stability of steady finite amplitude convection, *J. Fluid Mech.* **23**, 129–44.
Schrödingei, E (1926a) Quantisierung als Eigenertproblem (Part 1), *Annalen der Physik* **79**, 361–76
(1926b) Über das Verhältnis der Heisenberg–Born–Joidanschen Quantenmechanik zur der meinen, *Annalen der Physik* **79**, 734–55.
(1926c) Quantisierung als Eigenertproblem (Part 3) *Annalen der Physik* **80**, 437–90
(1926d) Quantisierung als Eigenertproblem (Part 4) *Annalen der Physik* **81**, 109–39.
Segel, L. A (1969) Distant side-walls cause slow amplitude modulation of cellular convection, *J Fluid Mech* **38**, 203–24
Shechtman, D , Blech, I., Gratias, D and Cahn, J W (1984) Metallic phase with long-range orientational order and no translational symmetry, *Phy Rev. Lett.* **53**, 1951–3.
Siggia, E D and Zippelius, A. (1981) Dynamics of defects in Rayleigh–Bénard convection, *Phys Rev. A* **24**, 1036–49
Silber, M and Proctor, M. R E (1998) Nonlinear competition between small and large hexagonal patterns, *Phys. Rev. Lett* **81**, 2450–3.
Sivashinsky, G. I. (1977) Nonlinear analysis of hydrodynamic instability in laminar flames Part I Derivation of basic equations, *Acta Astronautica* **4**, 1177–206
Skinner, G S and Swinney, H L. (1991) Periodic to quasiperiodic transition of chemical spiral rotation, *Physica D* **48**, 1–16.
Steeb, W.-H. (1996) *Continuous Symmetries, Lie Algebras, Differential Equations and Computer Algebra* (Singapore: World Scientific)
Sushchik, M. M. and Tsimring, L S (1994) The Eckhaus instability in hexagonal patterns, *Physica D* **74**, 90–106.
Swift, J and Hohenberg, P. C. (1977) Hydrodynamic fluctuations at the convective instability, *Phys. Rev. A* **15**, 319–28.
Tesauro, G and Cross, M. C. (1986) Climbing of dislocations in nonequilibrium patterns, *Phys Rev. A* **34**, 1363–79
(1987) Grain boundaries in models of convective patterns, *Philos Mag A* **56**, 703–24
Tse, D. P, Rucklidge, A. M , Hoyle, R B. and Silber, M (2000) Spatial period-multiplying instabilities of hexagonal Faraday waves, *Physica D.* **146**, 367–87
Tuckerman, L. S. and Barkley, D (1990) Bifurcation analysis of the Eckhaus instability, *Physica D* **46**, 57–86.

Turing, A M (1952) The chemical basis of morphogenesis, *Phil Trans R Soc Lond B* **237**, 37–72
Tveitereid, M. and Palm, E (1976) Convection due to internal heat sources, *J. Fluid Mech.* **76**, 481–99
Vanderbauwhede, A. (1980) Local bifurcation and symmetry, Habilitation thesis (Gent: Rijksuniversiteit)
Vasiev, B., Siegert, F and Weijer, C. (1997) Multiarmed spirals in excitable media, *Phys. Rev. Lett* **78**, 2489–92.
Walter, Th., Pesch, W and Bodenschatz, E. (2004) Dislocation dynamics in Rayleigh–Bénard convection, *Chaos* **14**, 933–9
White, D B. (1988) The planforms and onset of convection with a temperature-dependent viscosity, *J. Fluid Mech.* **191**, 247–86
Willis, G. E and Deardorff, J. W (1970) The oscillatory motions of Rayleigh convection, *J. Fluid Mech.* **44**, 661–72.
Winfree, A T. (1973) Scroll-shaped waves of chemical activity in three dimensions, *Science* **181**, 937–9
(1998) Evolving perspectives during 12 years of electrical turbulence, *Chaos* **8**, 1–19
Wulff, C. (2000) Transitions from relative equilibria to relative periodic orbits. *Documenta Math* **5**, 227–74.
(2002) Spiral waves and Euclidean symmetries. *Z Phys Chem* **216**, 535–50.
Yang, L., Dolnik, M., Zhabotinsky, A. M. and Epstein, I. R. (2002) Spatial resonances and superposition patterns in a reaction-diffusion model with interacting Turing modes, *Phys Rev. Lett.* **88**, 208303
Zaikin, A and Zhabotinsky, A. M (1970) Concentration wave propagation in a two-dimensional liquid-phase self-oscillating system, *Nature* **225**, 535–7.
Zhou, L. Q. and Ouyang, Q (2000) Experimental studies on long-wavelength instability and spiral breakup in a reaction-diffusion system, *Phys Rev. Lett* **85**, 1650–3
Zippelius, A. and Siggia, E. D (1983) Stability of finite-amplitude convection, *Phys Fluids* **26**, 2905–15

Index