

Nonlinear Dynamics

A Primer

ALFREDO MEDIO

MARJI LINES



CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014, Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

<http://www.cambridge.org>

© Alfredo Medio and Marji Lines 2001

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2001

Printed in the United Kingdom at the University Press, Cambridge

Typeface Computer Modern 11/14pt. *System* L^AT_EX [DBD]

A catalogue record of this book is available from the British Library

ISBN 0 521 55186 2 hardback
ISBN 0 521 55874 3 paperback

Contents

| | |
|--|----------------|
| <i>Preface</i> | <i>page xi</i> |
| 1 Statics and dynamics: some elementary concepts | 1 |
| 1.1 A static problem | 1 |
| 1.2 A discrete-time dynamic problem | 3 |
| 1.3 A continuous-time dynamic problem | 7 |
| 1.4 Flows and maps | 11 |
| Exercises | 19 |
| 2 Review of linear systems | 22 |
| 2.1 Introduction | 22 |
| 2.2 General solutions of linear systems in continuous time | 25 |
| 2.3 Continuous systems in the plane | 34 |
| 2.4 General solutions of linear systems in discrete time | 43 |
| 2.5 Discrete systems in the plane | 47 |
| 2.6 An economic example | 52 |
| Appendix: phase diagrams | 56 |
| Exercises | 61 |
| 3 Stability of fixed points | 66 |
| 3.1 Some formal definitions of stability | 66 |
| 3.2 The linear approximation | 71 |
| 3.3 The second or direct method of Lyapunov | 76 |
| Appendix A: general economic equilibrium | 85 |
| Appendix B: optimal economic growth | 92 |
| Appendix C: manifolds and tangent spaces | 98 |
| Exercises | 99 |

| | | |
|----------|---|-----|
| 4 | Invariant and attracting sets, periodic and quasiperiodic orbits | 104 |
| 4.1 | Invariant and limit sets | 105 |
| 4.2 | Stability and attractiveness of general sets | 110 |
| 4.3 | Attracting sets and attractors | 114 |
| 4.4 | Periodic orbits and their stability | 118 |
| 4.5 | Quasiperiodic orbits | 126 |
| | Appendix: conservative and dissipative systems | 128 |
| | Exercises | 130 |
| 5 | Local bifurcations | 133 |
| 5.1 | Introduction | 133 |
| 5.2 | Centre manifold theory | 134 |
| 5.3 | Local bifurcations for flows | 139 |
| 5.4 | Local bifurcations for maps | 151 |
| 5.5 | Bifurcations in two-dimensional maps | 160 |
| | Exercises | 161 |
| 6 | Chaotic sets and chaotic attractors | 163 |
| 6.1 | Basic definitions | 164 |
| 6.2 | Symbolic dynamics and the shift map | 166 |
| 6.3 | Logistic map with $\mu > 2 + \sqrt{5}$ | 171 |
| 6.4 | Smale horseshoe | 173 |
| 6.5 | Tent map and logistic map | 176 |
| 6.6 | Doubling maps | 178 |
| 6.7 | Chaotic attractors | 180 |
| 6.8 | The Lorenz model | 182 |
| | Exercises | 189 |
| 7 | Characteristic exponents, fractals, homoclinic orbits | 193 |
| 7.1 | Lyapunov characteristic exponents | 193 |
| 7.2 | Fractal dimension | 198 |
| 7.3 | Horseshoes and homoclinic orbits | 204 |
| | Exercises | 211 |
| 8 | Transition to chaos | 214 |
| 8.1 | Period-doubling route to chaos | 214 |
| 8.2 | Intermittency | 222 |
| 8.3 | Crises | 226 |
| 8.4 | Quasiperiodicity and chaos | 232 |
| | Exercises | 235 |

| | | |
|-----------|---|-----|
| 9 | The ergodic approach | 237 |
| 9.1 | Ergodic measures | 238 |
| 9.2 | Lyapunov characteristic exponents revisited | 247 |
| 9.3 | Natural, absolutely continuous, SBR measures | 249 |
| 9.4 | Attractors as invariant measures | 252 |
| 9.5 | Predictability, entropy | 253 |
| 9.6 | Isomorphism | 260 |
| 9.7 | Aperiodic and chaotic dynamics | 261 |
| 9.8 | Mixing | 265 |
| | Appendix: Shannon's entropy and Khinchin's axioms | 267 |
| | Exercises | 268 |
| 10 | Deterministic systems and stochastic processes | 270 |
| 10.1 | Bernoulli dynamics | 270 |
| 10.2 | Markov shifts | 276 |
| 10.3 | α -congruence | 277 |
| | <i>Further reading</i> | 282 |
| | <i>Bibliography</i> | 287 |
| | <i>Subject index</i> | 295 |

1

Statics and dynamics: some elementary concepts

Dynamics is the study of the movement through time of variables such as heartbeat, temperature, species population, voltage, production, employment, prices and so forth.

This is often achieved by means of equations linking the values of variables at different, uniformly spaced instants of time, i.e., **difference equations**, or by systems relating the values of variables to their time derivatives, i.e., **ordinary differential equations**. Dynamical phenomena can also be investigated by other types of mathematical representations, such as partial differential equations, lattice maps or cellular automata. In this book, however, we shall concentrate on the study of systems of difference and differential equations and their dynamical behaviour.

In the following chapters we shall occasionally use models drawn from economics to illustrate the main concepts and methods. However, in general, the mathematical properties of equations will be discussed independently of their applications.

1.1 A static problem

To provide a first, broad idea of the problems posed by dynamic *vis-à-vis* static analysis, we shall now introduce an elementary model that could be labelled as ‘supply-demand-price interaction in a single market’. Our model considers the quantities supplied and demanded of a single good, defined as functions of a single variable, its price, p . In economic parlance, this would be called partial analysis since the effect of prices and quantities determined in the markets of all other goods is neglected. It is assumed that the demand function $D(p)$ is decreasing in p (the lower the price, the greater the amount that people wish to buy), while the supply function $S(p)$ is increasing in p (the higher the price, the greater the amount that people wish to supply).

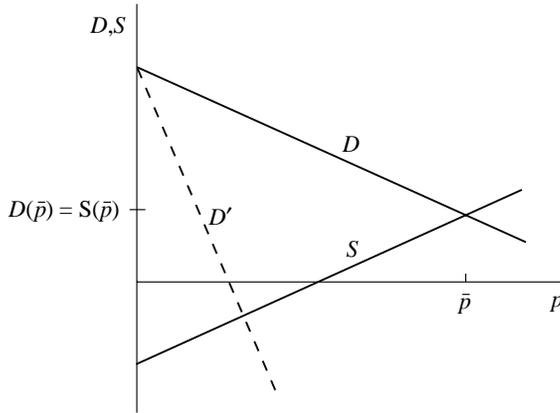


Fig. 1.1 The static partial equilibrium model

For example, in the simpler, linear case, we have:

$$\begin{aligned} D(p) &= a - bp \\ S(p) &= -m + sp \end{aligned} \quad (1.1)$$

and a, b, m and s are positive constants. Only nonnegative values of these variables are economically meaningful, thus we only consider $D, S, p \geq 0$. The **economic equilibrium condition** requires that the market of the good clears, that is demand equals supply, namely:

$$D(p) = S(p) \quad (1.2)$$

or

$$a - bp = -m + sp.$$

STATIC SOLUTION Mathematically, the solution to our problem is the value of the variable that solves (1.2) (in this particular case, a linear equation). Solving (1.2) for p we find:

$$\bar{p} = \frac{a + m}{b + s}$$

where \bar{p} is usually called the **equilibrium price** (see figure 1.1).¹ We call the problem *static* since no reference is made to time or, if you prefer,

¹The demand curve D' in figure 1.1 is provided to make the point that, with no further constraints on parameter values, the equilibrium price could imply negative equilibrium quantities of supply and demand. To eliminate this possibility we further assume that $0 < m/s \leq a/b$, as is the case for the demand curve D .

everything happens at the same time. Notice that, even though the static model allows us to find the equilibrium price of the good, it tells us nothing about what happens if the actual price is different from its equilibrium value.

1.2 A discrete-time dynamic problem

The introduction of dynamics into the model requires that we replace the equilibrium condition (1.2) with some hypothesis concerning the behaviour of the system off-equilibrium, i.e., when demand and supply are not equal. For this purpose, we assume the most obvious mechanism of price adjustment: over a certain interval of time, the price increases or decreases in proportion to the excess of demand over supply, $(D - S)$ (for short, **excess demand**). Of course, excess demand can be a positive or a negative quantity. Unless the adjustment is assumed to be instantaneous, prices must now be dated and p_n denotes the price of the good at time n , time being measured at equal intervals of length h . Formally, we have

$$p_{n+h} = p_n + h\theta[D(p_n) - S(p_n)]. \quad (1.3)$$

Since h is the period of time over which the adjustment takes place, θ can be taken as a measure of the speed of price response to excess demand. For simplicity, let us choose $h = 1$, $\theta = 1$. Then we have, making use of the demand and supply functions (1.1),

$$p_{n+1} = a + m + (1 - b - s)p_n. \quad (1.4)$$

In general, a solution of (1.4) is a *function of time* $p(n)$ (with n taking discrete, integer values) that satisfies (1.4).²

DYNAMIC SOLUTION To obtain the full dynamic solution of (1.4), we begin by setting $\alpha = a + m$, $\beta = (1 - b - s)$ to obtain

$$p_{n+1} = \alpha + \beta p_n. \quad (1.5)$$

To solve (1.5), we first set it in a canonical form, with all time-referenced terms of the variable on the left hand side (LHS), and all constants on the right hand side (RHS), thus:

$$p_{n+1} - \beta p_n = \alpha. \quad (1.6)$$

Then we proceed in steps as follows.

²We use the forms p_n and $p(n)$ interchangeably, choosing the latter whenever we prefer to emphasise that p is a function of n .

STEP 1 We solve the **homogeneous equation**, which is formed by setting the RHS of (1.6) equal to 0, namely:

$$p_{n+1} - \beta p_n = 0. \quad (1.7)$$

It is easy to see that a function of time $p(n)$ satisfying (1.7) is $p(n) = C\beta^n$, with C an arbitrary constant. Indeed, substituting in (1.7), we have

$$C\beta^{n+1} - \beta C\beta^n = C\beta^{n+1} - C\beta^{n+1} = 0.$$

STEP 2 We find a **particular solution** of (1.6), assuming that it has a form similar to the RHS in the general form. Since the latter is a constant, set $p(n) = k$, k a constant, and substitute it into (1.6), obtaining

$$k - \beta k = \alpha$$

so that

$$k = \frac{\alpha}{1 - \beta} = \frac{a + m}{b + s} = \bar{p} \quad \text{again!}$$

It follows that the $p(n) = \bar{p}$ is a solution to (1.6) and the constant (or stationary) solution of the dynamic problem is simply the solution of the static problem of section 1.1.

STEP 3 Since (1.6) is linear, the sum of the homogeneous and the particular solution is again a solution,³ called the **general solution**. This can be written as

$$p(n) = \bar{p} + C\beta^n. \quad (1.8)$$

The arbitrary constant C can now be expressed in terms of the initial condition. Putting $p(0) \equiv p_0$, and solving (1.8) for C we have

$$p_0 = \bar{p} + C\beta^0 = \bar{p} + C$$

whence $C = p_0 - \bar{p}$, that is, the difference between the initial and equilibrium values of p . The general solution can now be re-written as

$$p(n) = \bar{p} + (p_0 - \bar{p})\beta^n. \quad (1.9)$$

Letting n take integer values $1, 2, \dots$, from (1.9) we can generate a sequence of values of p , a ‘history’ of that variable (and consequently, a history of quantities demanded and supplied at the various prices), once its value at any arbitrary instant of time is given. Notice that, since the function $p_{n+1} =$

³This is called the *superposition principle* and is discussed in detail in chapter 2 section 2.1.

$f(p_n)$ is **invertible**, i.e., the function f^{-1} is well defined, $p_{n-1} = f^{-1}(p_n)$ also describes the past history of p .

The value of p at each instant of time is equal to the sum of the equilibrium value (the solution to the static problem which is also the particular, stationary solution) and the initial disequilibrium ($p_0 - \bar{p}$), amplified or dampened by a factor β^n . There are therefore two basic cases:

- (i) $|\beta| > 1$. Any nonzero deviation from equilibrium is amplified in time, the equilibrium solution is unstable and as $n \rightarrow +\infty$, p_n asymptotically tends to $+\infty$ or $-\infty$.
- (ii) $|\beta| < 1$. Any nonzero deviation is asymptotically reduced to zero, $p_n \rightarrow \bar{p}$ as $n \rightarrow +\infty$ and the equilibrium solution is consequently stable.

First-order, discrete-time equations (where the order is determined as the difference between the extreme time indices) can also have fluctuating behaviour, called **improper oscillations**,⁴ owing to the fact that if $\beta < 0$, β^n will be positive or negative according to whether n is even or odd. Thus the sign of the adjusting component of the solution, the second term of the RHS of (1.9), oscillates accordingly. Improper oscillations are dampened if $\beta > -1$ and explosive if $\beta < -1$.

In figure 1.2 we have two representations of the motion of p through time. In figure 1.2(a) we have a line defined by the solution equation (1.5), and the bisector passing through the origin which satisfies the equation $p_{n+1} = p_n$. The intersection of the two lines corresponds to the constant, equilibrium solution. To describe the off-equilibrium dynamics of p , we start on the abscissa from an initial value $p_0 \neq \bar{p}$. To find p_1 , we move vertically to the solution line and sidewise horizontally to the ordinate. To find p_2 , we first reflect the value of p_1 by moving horizontally to the bisector and then vertically to the abscissa. From the point p_1 , we repeat the procedure proposed for p_0 (up to the solution line, left to the ordinate), and so on and so forth. The procedure can be simplified by omitting the intermediate step and simply moving up to the solution line and sidewise to the bisector, up again, and so on, as indicated in figure 1.2(a). It is obvious that for $|\beta| < 1$, at each iteration of the procedure the initial deviation from equilibrium is diminished again, see figure 1.2(b). For example, if $\beta = 0.7$, we have $\beta^2 = 0.49$, $\beta^3 = 0.34, \dots, \beta^{10} \approx 0.03, \dots$) and the equilibrium solution is approached asymptotically.

The reader will notice that stability of the system and the possibility

⁴The term *improper* refers to the fact that in this case oscillations of variables have a ‘kinky’ form that does not properly describe the smoother ups and downs of real variables. We discuss *proper* oscillations in chapter 3.

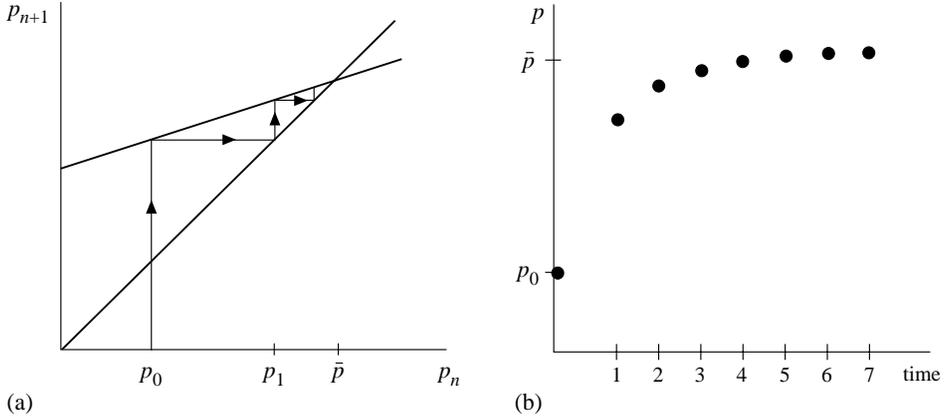


Fig. 1.2 Convergence to \bar{p} in the discrete-time partial equilibrium model

of oscillatory behaviour depends entirely on β and therefore on the two parameters b and s , these latter denoting respectively the slopes of the demand and supply curves. The other two parameters of the system, a and m , determine α and consequently they affect only the equilibrium value \bar{p} . We can therefore completely describe the dynamic characteristics of the solution (1.9) over the parameter space (b, s) . The boundary between stable and unstable behaviour is given by $|\beta| = 1$, and convergence to equilibrium is guaranteed for

$$\begin{aligned} -1 < \beta < 1 \\ 2 > b + s > 0. \end{aligned}$$

The assumptions on the demand and supply functions imply that $b, s > 0$. Therefore, the stability condition is $(b + s) < 2$, the stability boundary is the line $(b + s) = 2$, as represented in figure 1.3. Next, we define the curve $\beta = 1 - (b + s) = 0$, separating the zone of monotonic behaviour from that of improper oscillations, which is also represented in figure 1.3. Three zones are labelled according to the different types of dynamic behaviour, namely: convergent and monotonic; convergent and oscillatory; divergent and oscillatory. Since $b, s > 0$, we never have the case $\beta > 1$, corresponding to divergent, nonoscillatory behaviour.

If $|\beta| > 1$ any initial difference $(p_0 - \bar{p})$ is amplified at each step. In this model, we can have $|\beta| > 1$ if and only if $\beta < -1$. Instability, then, is due to **overshooting**. Any time the actual price is, say, too low and there is positive excess demand, the adjustment mechanism generates a change in the price in the ‘right’ direction (the price rises) but the change is too large.

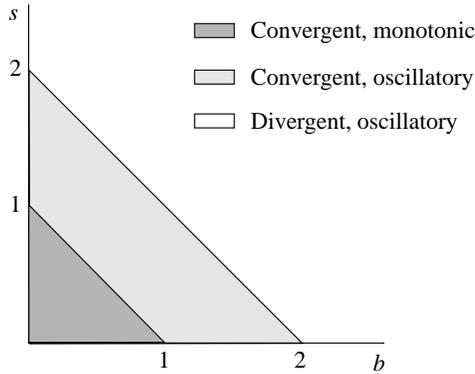


Fig. 1.3 Parameter space for the discrete-time partial equilibrium model

After the correction, the new price is too high (negative excess demand) and the discrepancy from equilibrium is larger than before. A second adjustment follows, leading to another price that is far too low, and so on. We leave further study of this case to the exercises at the end of this chapter.

1.3 A continuous-time dynamic problem

We now discuss our simple dynamical model in a continuous-time setting. Let us consider, again, the price adjustment equation (1.3) (with $\theta = 1$, $h > 0$) and let us adopt the notation $p(n)$ so that

$$p(n+h) = p(n) + h(D[p(n)] - S[p(n)]).$$

Dividing this equation throughout by h , we obtain

$$\frac{p(n+h) - p(n)}{h} = D[p(n)] - S[p(n)]$$

whence, taking the limit of the LHS as $h \rightarrow 0$, and recalling the definition of a derivative, we can write

$$\frac{dp(n)}{dn} = D[p(n)] - S[p(n)].$$

Taking the interval h to zero is tantamount to postulating that time is a continuous variable. To signal that time is being modelled differently we substitute the time variable $n \in \mathbb{Z}$ with $t \in \mathbb{R}$ and denote the value of p at time t simply by p , using the extended form $p(t)$ when we want to emphasise that price is a function of time. We also make use of the efficient Newtonian

notation $dx(t)/dt = \dot{x}$ to write the price adjustment mechanism as

$$\frac{dp}{dt} = \dot{p} = D(p) - S(p) = (a + m) - (b + s)p. \quad (1.10)$$

Equation (1.10) is an ordinary differential equation relating the values of the variable p at a given time t to its first derivative with respect to time at the same moment. It is **ordinary** because the solution $p(t)$ is a function of a single independent variable, time. Partial differential equations, whose solutions are functions of more than one independent variable, will not be treated in this book, and when we refer to differential equations we mean ordinary differential equations.

DYNAMIC SOLUTION The dynamic problem is once again that of finding a function of time $p(t)$ such that (1.10) is satisfied for an arbitrary initial condition $p(0) \equiv p_0$.

As in the discrete-time case, we begin by setting the equation in canonical form, with all terms involving the variable or its time derivatives on the LHS, and all constants or functions of time (if they exist) on the RHS, thus

$$\dot{p} + (b + s)p = a + m. \quad (1.11)$$

Then we proceed in steps as follows.

STEP 1 We solve the homogeneous equation, formed by setting the RHS of (1.11) equal to 0, and obtain

$$\dot{p} + (b + s)p = 0 \text{ or } \dot{p} = -(b + s)p. \quad (1.12)$$

If we now integrate (1.12) by separating variables, we have

$$\int \frac{dp}{p} = -(b + s) \int dt$$

whence

$$\ln p(t) = -(b + s)t + A$$

where A is an arbitrary integration constant. Taking now the antilogarithm of both sides and setting $e^A = C$, we obtain

$$p(t) = Ce^{-(b+s)t}.$$

STEP 2 We look for a particular solution to the nonhomogeneous equation (1.11). The RHS is a constant so we try $p = k$, k a constant and consequently $\dot{p} = 0$. Therefore, we have

$$\dot{p} = 0 = (a + m) - (b + s)k$$

whence

$$k = \frac{a + m}{b + s} = \bar{p}.$$

Once again the solution to the static problem turns out to be a special (stationary) solution to the corresponding dynamic problem.

STEP 3 Since (1.12) is linear, the general solution can be found by summing the particular solution and the solution to the homogeneous equation, thus

$$p(t) = \bar{p} + Ce^{-(b+s)t}.$$

Solving for C in terms of the initial condition, we find

$$p(0) \equiv p_0 = \bar{p} + C \text{ and } C = (p_0 - \bar{p}).$$

Finally, the complete solution to (1.10) in terms of time, parameters, initial and equilibrium values is

$$p(t) = \bar{p} + (p_0 - \bar{p})e^{-(b+s)t}. \quad (1.13)$$

As in the discrete case, the solution (1.13) can be interpreted as the sum of the equilibrium value and the initial deviation of the price variable from equilibrium, amplified or dampened by the term $e^{-(b+s)t}$. Notice that in the continuous-time case, a solution to a differential equation $\dot{p} = f(p)$ always determines both the future and the past history of the variable p , independently of whether the function f is invertible or not. In general, we can have two main cases, namely:

- (i) $(b + s) > 0$ Deviations from equilibrium tend asymptotically to zero as $t \rightarrow +\infty$.
- (ii) $(b + s) < 0$ Deviations become indefinitely large as $t \rightarrow +\infty$ (or, equivalently, deviations tend to zero as $t \rightarrow -\infty$).

Given the assumptions on the demand and supply functions, and therefore on b and s , the explosive case is excluded for this model. If the initial price is below its equilibrium value, the adjustment process ensures that the price increases towards it, if the initial price is above equilibrium, the price declines to it. (There can be no overshooting in the continuous-time case.) In a manner analogous to the procedure for difference equations, the equilibria

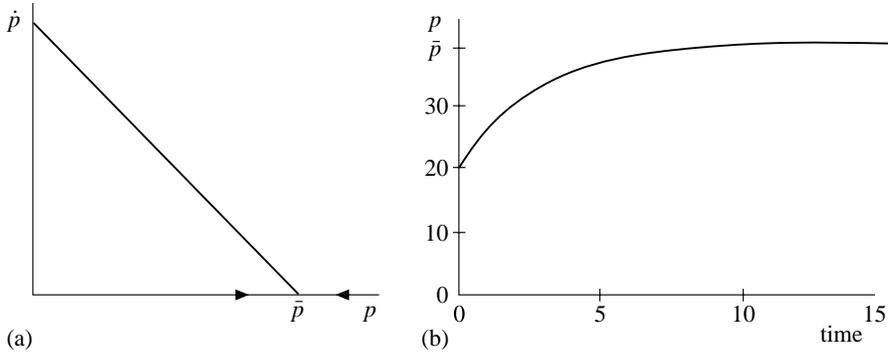


Fig. 1.4 The continuous-time partial equilibrium model

of differential equations can be determined graphically in the plane (p, \dot{p}) as suggested in figure 1.4(a). Equilibria are found at points of intersection of the line defined by (1.10) and the abscissa, where $\dot{p} = 0$. Convergence to equilibrium from an initial value different from the equilibrium value is shown in figure 1.4(b).

Is convergence likely for more general economic models of price adjustment, where other goods and income as well as substitution effects are taken into consideration? A comprehensive discussion of these and other related microeconomic issues is out of the question in this book. However, in the appendixes to chapter 3, which are devoted to a more systematic study of stability in economic models, we shall take up again the question of convergence to or divergence from economic equilibrium.

We would like to emphasise once again the difference between the discrete-time and the continuous-time formulation of a seemingly identical problem, represented by the two equations

$$p_{n+1} - p_n = (a + m) - (b + s)p_n \quad (1.4)$$

$$\dot{p} = (a + m) - (b + s)p. \quad (1.10)$$

Whereas in the latter case $(b+s) > 0$ is a sufficient (and necessary) condition for convergence to equilibrium, stability of (1.4) requires that $0 < (b+s) < 2$, a tighter condition.

This simple fact should make the reader aware that a naive translation of a model from discrete to continuous time or vice versa may have unsuspected consequences for the dynamical behaviour of the solutions.

1.4 Flows and maps

To move from the elementary ideas and examples considered so far to a more general and systematic treatment of the matter, we need an appropriate mathematical framework, which we introduce in this section. When necessary, the most important ideas and methods will be discussed in greater detail in the following chapters. For the sake of presentation, we shall begin with continuous-time systems of differential equations, which typically take the canonical form

$$\frac{dx}{dt} = \dot{x} = f(x) \tag{1.14}$$

where f is a function with domain U , an open subset of \mathbb{R}^m , and range \mathbb{R}^m (denoted by $f:U \rightarrow \mathbb{R}^m$). The vector⁵ $x = (x_1, x_2, \dots, x_m)^T$ denotes the physical variables to be studied, or some appropriate transformations of them; $t \in \mathbb{R}$ indicates time. The variables x_i are sometimes called ‘dependent variables’ whereas t is called the ‘independent variable’.

Equation (1.14) is called **autonomous** when the function f does not depend on time directly, but only through the state variable x . In this book we shall be mainly concerned with this type of equation, but in our discussions of stability in chapters 3 and 4 we shall have something to say about nonautonomous equations as well.

The space \mathbb{R}^m , or an appropriate subspace of dependent variables — that is, variables whose values specify the state of the system — is referred to as the **state space**. It is also known as the **phase space** or, sometimes, the **configuration space**, but we will use only the first term. Although for most of the problems encountered in this book the state space is the Euclidean space, we occasionally discuss dynamical systems different from \mathbb{R}^m , such as the unit circle. The circle is a one-dimensional object embedded in a two-dimensional Euclidean space. It is perhaps the simplest example of a kind of set called **manifold**. Roughly speaking, a manifold is a set which locally looks like a piece of \mathbb{R}^m . A more precise definition is deferred to appendix C of chapter 3, p. 98.

In simple, low-dimensional graphical representations of the state space the direction of motion through time is usually indicated by arrows pointing to the future. The enlarged space in which the time variable is explicitly

⁵Recall that the **transposition operator**, or transpose, designated by T , when applied to a row vector, returns a column vector and vice versa. When applied to a matrix, the operator interchanges rows and columns. Unless otherwise indicated, vectors are column vectors.

considered is called the **space of motions**. Schematically, we have

$$\begin{array}{ccccc} \mathbb{R} & \times & \mathbb{R}^m & = & \mathbb{R}^{1+m} \\ \downarrow & & \downarrow & & \downarrow \\ \text{time} & & \text{state} & & \text{space of} \\ & & \text{space} & & \text{motions} \end{array}$$

The function f defining the differential equation (1.14) is also called a **vector field**, because it assigns to each point $x \in U$ a velocity vector $f(x)$. A solution of (1.14) is often written as a function $x(t)$, where $x : I \rightarrow \mathbb{R}^m$ and I is an interval of \mathbb{R} . If we want to specifically emphasise the solution that, at the initial time t_0 , passes through the initial point x_0 , we can write $x(t; t_0, x_0)$, where $x(t_0; t_0, x_0) = x(t_0) = x_0$. We follow the practice of setting $t_0 = 0$ when dealing with autonomous systems whose dynamical properties do not depend on the choice of initial time.

remark 1.1 In applications, we sometimes encounter differential equations of the form

$$\frac{d^m x}{dt^m} = F\left(x, \frac{dx}{dt}, \dots, \frac{d^{m-1}x}{dt^{m-1}}\right) \quad x \in \mathbb{R} \quad (1.15)$$

where $d^k x/dt^k$ denotes the k th derivative of x with respect to time. Equation (1.15) is an autonomous, ordinary differential equation of order m , where m is the highest order of differentiation with respect to time appearing in the equation. It can always be put into the canonical form (1.14) by introducing an appropriate number of auxiliary variables. Specifically, put

$$\frac{d^k x}{dt^k} = z_{k+1}, \quad 0 \leq k \leq m-1$$

(where, for $k=0$, $d^k x/dt^k = x$) so that

$$\begin{aligned} \dot{z}_k &= z_{k+1}, & 1 \leq k \leq m-1 \\ \dot{z}_m &= F(z_1, \dots, z_m). \end{aligned}$$

If we now denote by $z \in \mathbb{R}^m$ the m -dimensional vector $(z_1, \dots, z_m)^T$ we can write:

$$\dot{z} = f(z)$$

where $f(z) = [z_2, \dots, z_m, F(z_1, \dots, z_m)]^T$.

We can also think of solutions of differential equations in a different manner which is now prevalent in dynamical system theory and will be very helpful for understanding some of the concepts discussed in the following chapters.

If we denote by $\phi_t(x) = \phi(t, x)$ the state in \mathbb{R}^m reached by the system at time t starting from x , then the totality of solutions of (1.14) can be represented by a one-parameter family of maps⁶ $\phi_t : U \rightarrow \mathbb{R}^m$ satisfying

$$\left. \frac{d}{dt}[\phi(t, x)] \right|_{t=\tau} = f[\phi(\tau, x)]$$

for all $x \in U$ and for all $\tau \in I$ for which the solution is defined.

The family of maps $\phi_t(x) = \phi(t, x)$ is called the flow (or the flow map) generated by the vector field f . If f is continuously differentiable (that is, if all the functions in the vector are continuously differentiable), then for any point x_0 in the domain U there exists a $\delta(x_0) > 0$ such that the solution $\phi(t, x_0)$ through that point exists and is unique for $|t| < \delta$. The existence and uniqueness result is *local* in time in the sense that δ need not extend to (plus or minus) infinity and certain vector fields have solutions that ‘explode’ in finite time (see exercise 1.8(c) at the end of the chapter).

When the solution of a system of differential equations $\dot{x} = f(x)$ is not defined for all time, a new system $\dot{x} = g(x)$ can be determined which has the same forward and backward orbits *in the state space* and such that each orbit is defined for all time. If $\psi(t, x)$ is the flow generated by the vector field g , the relation between ψ and the flow ϕ generated by f is the following:

$$\psi(t, x) = \phi[\tau(t, x), x] \quad x \in U$$

and

$$\tau : \mathbb{R} \times U \rightarrow \mathbb{R}$$

is a time-reparametrisation function monotonically increasing in t for all $x \in U$.

EXAMPLE Suppose we have a system

$$\dot{x} = f(x) \tag{1.16}$$

with $f : \mathbb{R}^m \rightarrow \mathbb{R}^m$, a continuously differentiable function with flow $\phi(t, x)$ defined on a maximal time interval $-\infty < a < 0 < b < +\infty$. Then the

⁶The terms map or mapping indicate a function. In this case, we speak of $y = f(x)$ as the **image** of x under the map f . If f is invertible, we can define the inverse function f^{-1} , that is, the function satisfying $f^{-1}[f(x)] = x$ for all x in the domain of f and $f[f^{-1}(y)] = y$ for all y in the domain of f^{-1} . Even if f is not invertible, the notation $f^{-1}(y)$ makes sense: it is the **set of pre-images** of y , that is, all points x such that $f(x) = y$. The terms map, mapping are especially common in the theory of dynamical systems where iterates of a map are used to describe the evolution of a variable in time.

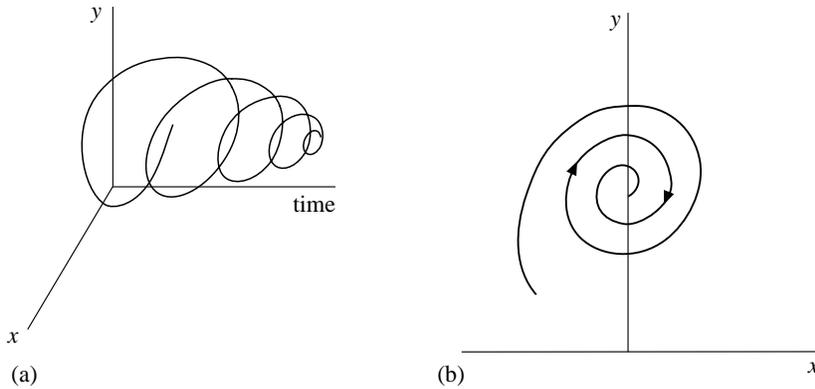


Fig. 1.5 A damped oscillator in \mathbb{R}^2 : (a) space of motions; (b) state space

differential equation $\dot{x} = g(x)$ with $g(x) : \mathbb{R}^m \rightarrow \mathbb{R}^m$ and

$$g(x) = \frac{f(x)}{1 + \|f(x)\|}$$

(where $\|\cdot\|$ denotes Euclidean norm), defines a dynamical system whose forward and backward orbits are the same as those of (1.16) but whose solutions are defined for all time.⁷

The set of points $\{\phi(t, x_0) \mid t \in I\}$ defines an **orbit** of (1.14), starting from a given point x_0 . It is a solution curve in the state space, parametrised by time. The set $\{[t, \phi(t, x_0)] \mid t \in I\}$ is a **trajectory** of (1.14) and it evolves in the space of motions. However, in applications, the terms orbit and trajectory are often used as synonyms. A simple example of a trajectory in the space of motions $\mathbb{R} \times \mathbb{R}^2$ and the corresponding orbit in the state space \mathbb{R}^2 is given in figure 1.5. Clearly the orbit is obtained by projecting the trajectory onto the state space.

The flows generated by vector fields form a very important subset of a more general class of maps, characterised by the following definition.

definition 1.1 A **flow** is a map $\phi : I \subset \mathbb{R} \times X \rightarrow X$ where X is a metric space, that is, a space endowed with a distance function, and ϕ has the following properties

- (a) $\phi(0, x) = x$ for every $x \in X$ (identity axiom);
- (b) $\phi(t + s, x) = \phi[s, \phi(t, x)] = \phi[t, \phi(s, x)] = \phi(s + t, x)$, that is, time-translated solutions remain solutions;

⁷For details see Bhatia and Szegö (1970), p. 78; Robinson (1999), pp. 146–7.

(c) for fixed t , ϕ_t is a homeomorphism on X .

Alternatively, and equivalently, a flow may be defined as a one-parameter family of maps $\phi_t : X \rightarrow X$ such that the properties (a)–(c) above hold for all $t, s \in \mathbb{R}$.

remark 1.2 A **distance** on a space X (or, a **metric** on X) is a function $X \times X \rightarrow \mathbb{R}^+$ satisfying the following properties for all $x, y \in X$:

- (1) $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$;
- (2) $d(x, y) = d(y, x)$ (symmetry);
- (3) $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality).

Notice that there also exist notions of distance which are perfectly meaningful but do not satisfy the definition above and therefore do not define a metric, for example:

the distance between a point and a set A ;

$$d(x, A) = \inf_{y \in A} d(x, y).$$

the distance between two sets A and B

$$d(A, B) = \inf_{x \in A} \inf_{y \in B} d(x, y).$$

Neither of these cases satisfies property (1) in remark 1.2. However, there exists a ‘true’ distance between two sets which is a metric in the space of nonempty, compact sets, i.e., the Hausdorff distance.⁸

In this book we are mainly concerned with applications for which ϕ is a flow generated by a system of differential equations and the state space is an Euclidean space or, sometimes, a manifold. However, some concepts and results in later chapters of the book will be formulated more generally in terms of flows on metric spaces.

Consider now a system of nonautonomous differential equations such that

$$\dot{x} = f(t, x) \tag{1.17}$$

where $f : \mathbb{R} \times U \rightarrow \mathbb{R}^m$, and assume that a unique solution exists for all $(t_0, x_0) \in \mathbb{R} \times U$. Then we can represent solutions of (1.17) by means of a flow $\phi : \mathbb{R} \times X \rightarrow X$, where $X \subset (\mathbb{R} \times \mathbb{R}^m)$. This suggests that a

⁸See, for example, Edgar (1990), pp. 65–6.

nonautonomous system $\dot{x} = f(t, x)$ can be transformed into an equivalent autonomous system by introducing an arbitrary variable $\theta = t$ and writing

$$\begin{aligned}\dot{\theta} &= 1 \\ \dot{x} &= f(\theta, x).\end{aligned}\tag{1.18}$$

Notice that, by definition, the extended autonomous system (1.18) has no equilibrium point in X . However, if the original, nonautonomous system (1.17) has a *uniformly* stable (*uniformly*, asymptotically stable) equilibrium point, then for the extended autonomous system (1.18), the t -axis is a stable (asymptotically stable) invariant set. The precise meaning of (asymptotic, uniform) stability will be discussed in chapters 3 and 4.

Solutions of system (1.14) can be written in either the simpler form $x(t)$, $x : I \rightarrow \mathbb{R}^m$, or $\phi_t(x) : U \rightarrow \mathbb{R}^m$, or again $\phi(t, x)$, $\phi : I \times U \rightarrow \mathbb{R}^m$, depending on what aspect of solutions one wants to emphasise. The notation $\phi_t(x)$ is especially suitable for discussing discrete-time maps derived from continuous-time systems.

If time t is allowed to take only uniformly sampled, discrete values, separated by a fixed interval τ , from a continuous-time flow we can derive a discrete-time map (a difference equation)

$$x_{n+\tau} = G(x_n)\tag{1.19}$$

where $G = \phi_\tau$. Certain properties of continuous-time dynamical systems are preserved by this transformation and can be studied by considering the discrete-time systems derived from them. If the unit of measure of time is chosen so that $\tau = 1$, we have the canonical form

$$x_{n+1} = G(x_n).\tag{1.20}$$

Let the symbol \circ denote the composition of functions, so that, $f \circ g(x)$ means $f[g(x)]$. Then we write

$$x_n = G(x_{n-1}) = G \circ G(x_{n-2}) = \dots = G \circ G \circ \dots \circ G(x_0) = G^n(x_0)$$

where G^n is the composition of G with itself n times, or the n th iteration of G , with $n \in \mathbb{Z}^+$. If G is invertible and G^{-1} is a well defined function, G^n with $n \in \mathbb{Z}^-$ denotes the n th iterate of G^{-1} . (Note that $G^n(x)$ is *not* the n th power of $G(x)$.) Thus, iterates of the map G (or G^{-1}) can be used to determine the value of the variable x at time n , when the initial condition x_0 is fixed.⁹

⁹For autonomous difference equations whose solutions do not depend on the choice of the initial time, in a manner analogous to our practice for autonomous differential equations, we take the initial time as zero.

remark 1.3 There exists another way of deriving a discrete-time map from a continuous-time dynamical system, called **Poincaré map**, which describes the sequence of positions of a system generated by the intersections of an orbit in continuous time and a given space with a lower dimension, called **surface of section**. Clearly, in this case the time intervals between different pairs of states of the systems need not be equal. Poincaré maps are a powerful method of investigation of dynamical systems and we shall make some use of them in chapter 4, when we discuss periodic solutions and in chapters 7 and 8.

Of course, there exist problems that are conceived from the beginning as discrete dynamical systems (difference equations). In fact, there are difference equations that cannot be derived from differential equations. In particular, this is true of *noninvertible maps* which have been extensively used in recent years in the study of dynamical problems in many applied fields. Intuitively, the reason why a noninvertible map cannot be a flow map (derived from a differential equation as explained above) is that such a map uniquely determines the dynamics in one time direction only whereas, under standard assumptions, solutions of a differential equation always determine the dynamics in both directions uniquely.

remark 1.4 Orbits of differential equations are continuous *curves*, while orbits of maps are *discrete sets of points*. This has a number of important consequences, the most important of which can be appreciated intuitively. If the solution of an autonomous system of differential equations through a point is unique, two solution curves cannot intersect one another in the state space. It follows that, for continuous-time dynamical systems of dimension one and two, the orbit structure must be drastically constrained. In the former, simpler case, we can only have fixed points and orbits leading to (or away from) them; in the two-dimensional case, nothing more complex than periodic orbits can occur. For maps the situation is different. It remains true that the orbit starting from a given point in space is uniquely determined in the direction defined by the map. However, since discrete-time orbits, so to speak, can ‘jump around’, even simple, one-dimensional nonlinear maps can generate very complicated orbits, as we shall see in the following chapters.

Generalising the simple examples discussed in sections 1.2 and 1.3 above, the stationary, equilibrium solutions of multi-dimensional dynamical systems in both continuous and discrete time can be identified by solving systems of equations.

In the former case, setting $\dot{x} = 0$ in (1.14) the set of **equilibrium** or

fixed points is defined by

$$E = \{\bar{x} | f(\bar{x}) = 0\}$$

that is, the set of values of x such that its rate of change in time is nil.

Analogously, in the discrete-time case,

$$x_{n+1} = G(x_n)$$

we have

$$E = \{\bar{x} | \bar{x} - G(\bar{x}) = 0\}$$

that is, the set of values of x that are mapped to themselves by G . Because the functions f and G are generally nonlinear, there are no ready-made procedures to find the equilibrium solutions exactly, although geometrical and numerical techniques often give us all the qualitative information we need. Notice that linear systems typically have a unique solution, whereas nonlinear systems typically have either no solutions, or a finite number of them. It follows that only nonlinear systems may describe the interesting phenomenon of (finite) multiple equilibria.

For a system of autonomous, differential equations like (1.14), a general solution $\phi(t, x)$ can seldom be written in a closed form, i.e., as a combination of known elementary functions (powers, exponentials, logarithms, sines, cosines, etc.). Unfortunately, closed-form solutions are available only for special cases, namely: systems of linear differential equations; one-dimensional differential equations (i.e., those for which $m = 1$); certain rather special classes of nonlinear differential equations of order greater than one (or systems of equations with $m > 1$). The generality of nonlinear systems which are studied in applications escapes full analytical investigation, that is to say, an exact mathematical description of solution orbits cannot be found. Analogous difficulties arise when dynamical systems are represented by means of nonlinear maps. In this case, too, closed-form solutions are generally available only for linear systems.

The importance of this point should not be exaggerated. On the one hand, even when a closed-form solution exists, it may not be very useful. A handbook of mathematical formulae will typically have a hundred pages of integrals for specific functions, so that a given nonlinear model may indeed have a solution. However, that solution may not provide much intuition, nor much information if the solution function is not a common, well known function. On the other hand, in many practical cases we are not especially interested in determining (or approximating) exact individual solutions, but we want to establish the qualitative properties of an ensemble

of orbits starting from certain practically relevant sets of initial conditions. These properties can often be investigated effectively by a combination of mathematical, geometrical, statistical and numerical methods. Much of what follows is dedicated precisely to the study of some of those methods.

Before turning to this goal, however, we review in chapter 2 the class of dynamical systems which is best understood: linear systems. Dynamical linear systems in both continuous and discrete time are not terribly interesting *per se* because their behaviour is morphologically rather limited and they cannot be used effectively to represent cyclical or complex dynamics. However, linear theory is an extremely useful tool in the analysis of nonlinear systems. For example, it can be employed to investigate qualitatively their **local** behaviour, e.g., their behaviour in a neighbourhood of a single point or of a periodic orbit. This is particularly important in stability analysis (chapters 3 and 4) and in the study of (local) bifurcations (chapter 5).

Exercises

- 1.1 Consider the discrete-time partial equilibrium model summarised in (1.6) given the parameter values $a = 10$, $b = 0.2$, $m = 2$, $s = 0.1$. Write the general solution given the initial values $p_0 = 20$ and $p_0 = 100$. Calculate the values for the price at time periods 0, 1, 2, 4, 10, 100 starting from each of the above initial values and sketch the trajectories for time periods 0–10.
- 1.2 State a parameter configuration for the discrete-time partial equilibrium model that implies $\beta < 0$. Describe the dynamics implied by that choice. Using these parameter values and $a = 10$, $m = 2$, sketch the dynamics in the space (p_n, p_{n+1}) . Draw the bisector line and from the chosen initial condition, iterate 3 or 4 times. Show the direction of movement with arrows.
- 1.3 If we define the parameters as in exercise 1.1 ($b = 0.2$, $s = 0.1$, $a = 10$, $m = 2$), the continuous-time, partial equilibrium model of (1.11) gives the constant exponent of the solution as $b + s = 0.3$. Let this be case 1. If $s = 0.6$, $b + s = 0.8$. Let this be case 2. Calculate the solution values for case 1 and case 2 at periods $t = 0, 1, 2, 4.67, 10, 100$ starting from the initial condition $p_0 = 20$. Comment on the speed of the adjustment process. Note the different integer values of t for which equilibrium in Case 2 is approximated using a precision of 1 decimal point, 2 decimal points, 3 decimal points.
- 1.4 Suppose that the good under consideration is a ‘Giffen’ good (for which $dD/dp > 0$ and therefore $b < 0$). It is unlikely, but possible

that $b + s < 0$. Sketch the differential equation (1.10) under that hypothesis in the (p, \dot{p}) plane, note the equilibrium point and comment on the adjustment process.

- 1.5 Convert these higher-order differential equations to systems of first-order differential equations and write the resulting systems in matrix form:

(a) $\ddot{x} + x = 1$
 (b) $\frac{d^3x}{dt^3} + 0.4\ddot{x} - 2x = 0$
 (c) $\frac{d^4x}{dt^4} + 4\ddot{x} - 0.5\dot{x} - x = 11$.

- 1.6 Convert the following higher-order system of differential equations into a system of first-order differential equations

$$\begin{aligned}\ddot{x} + x &= 1 \\ \ddot{y} - \dot{y} - y &= -1.\end{aligned}$$

- 1.7 Higher-order difference equations and systems can also be converted to first-order systems using auxiliary variables. A k th-order equation $x_{n+k} = G(x_{n+k-1}, \dots, x_n)$ can be converted by setting

$$\begin{aligned}x_n &= z_n^{(1)} \\ z_{n+1}^{(1)} &= x_{n+1} = z_n^{(2)} \\ z_{n+1}^{(2)} &= x_{n+2} = z_n^{(3)} \\ &\vdots \quad \quad \quad \vdots \\ z_{n+1}^{(k)} &= x_{n+k} = G(x_{n+k-1}, \dots, x_n) = G(z_n^{(k)}, \dots, z_n^{(1)}).\end{aligned}$$

Convert the following difference equations into systems of first-order difference equations and write them in matrix form

(a) $x_{n+2} - ax_{n+1} + bx_n = 1$
 (b) $0.5x_{n+3} + 2x_{n+1} - 0.1x_n = 2$.

- 1.8 Use integration techniques to find exact solutions to the following differential equations and sketch trajectories where possible, assuming an initial value of $x(0) = 1$

(a) $\dot{x} = 2x$
 (b) $\dot{x} = \frac{1}{x^2}$
 (c) $\dot{x} = x^2$.

- 1.9 Use the technique described in the example in section 1.4 to find a function g , defined over all time and such that $\dot{x} = g(x)$ has the same backward and forward orbits in the state space as $\dot{x} = x^2$.

- 1.10 Write the exact solution of the following differential equation (*Hint*: rewrite the equation as $dx/dt = \mu x(1 - x)$ and integrate, separating variables) and discuss the dynamics of x

$$\dot{x} = \mu x(1 - x) \quad x \in [0, 1].$$