[9c]

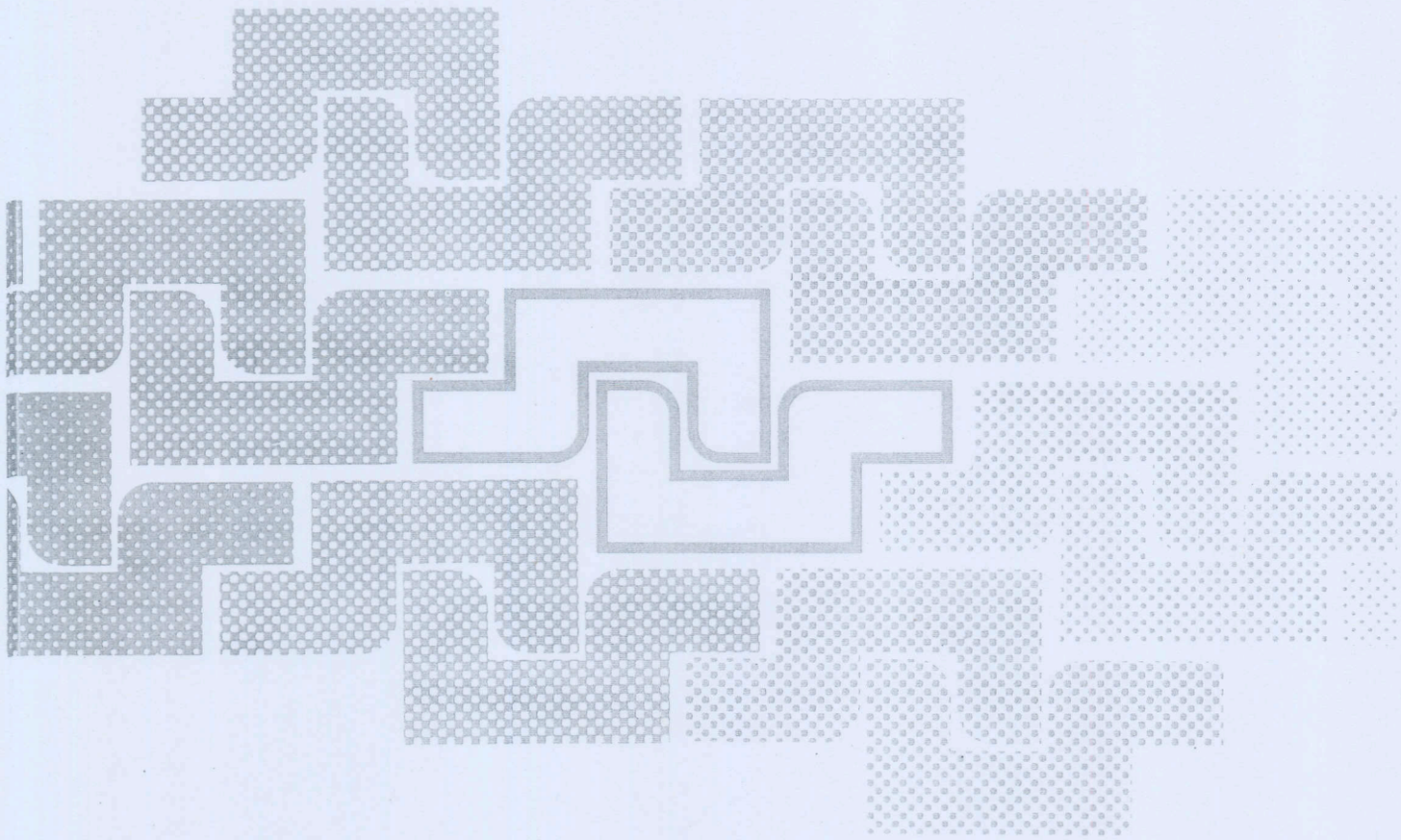
A.R. Humphries and A.M. Stuart,

Numerical Analysis of Deterministic and Random Dynamical Systems.

Appears in Modern *Methods in Scientific Computing and Applications,*

editors A. Bourlioux and M.J. Gander.

NATO Science Series II volume 75. Springer (2002).



Modern Methods in Scientific Computing and Applications

Edited by

Anne Bourlioux and Martin J. Gander

NATO Science Series

Deterministic and random dynamical systems: theory and numerics

A. R. HUMPHRIES

School of Mathematical Sciences
University of Sussex
Brighton BN1 9QH
England

A. M. STUART

Mathematics Institute
Warwick University
Coventry CV4 7AL
England

Abstract

The theory of (random) dynamical systems is a framework for the analysis of large time behaviour of time-evolving systems (driven by noise). These notes contain an elementary introduction to the theory of both dynamical and random dynamical systems. The subject matter is made accessible by means of very simple examples and highlights relationships between the deterministic and the random theories.

Throughout the notes both continuous and discrete time are considered. Continuous time problems are primarily ordinary and stochastic differential equations, whilst approximating numerical schemes are used to illustrate discrete time. The central role of Lyapunov functions is highlighted, both in inferring the existence of attractors in the deterministic case, and in the construction of invariant measures and random attractors in the random case. The importance of approximations which inherit Lyapunov functions is emphasised.

1 Deterministic dynamical systems

1.1 Differential equations

Let $f : \mathbb{R}^d \to \mathbb{R}^d$ be a continuous function. Consider the following ordinary differential equation (ODE)

$$\frac{dx}{dt} = f(x(t)), \quad x(0) = x_0. \tag{1.1.1}$$

Here the problem is to find $x \in C(\mathbb{R}^+, \mathbb{R}^d)$ which satisfies (1.1.1). We will often write $\dot{x}$ for dx/dt. We use the notation $\mathbb{R}^+ = [0, \infty)$ and we will later need $\mathbb{Z}^+ = \{0, 1, 2, 3, \dots\}$.

A. Bourlioux and M.J. Gander (eds.), Modern Methods in Scientific Computing and Applications, 211–254.

Equation (1.1.1) can also be written as an integral equation:

$$x(t) = x_0 + \int_0^t f(x(s))\,ds, \quad x(0) = x_0. \tag{1.1.2}$$

This formulation will be useful when we introduce stochastic differential equations.

1.1.1 Example (i) A scalar linear problem:

$$\dot{x} = \lambda x, \quad x(0) = x_0 \in \mathbb{R}. \tag{1.1.3}$$

This has exact solution $x(t) = e^{\lambda t} x_0$. Note that $x(t) \to 0$ as $t \to \infty$ if $\lambda < 0$. Since we know the exact solution this is a useful test problem for numerical algorithms.

(ii) A scalar nonlinear equation:

$$\dot{x} = f(x) = x - x^3, \quad x(0) = x_0. \tag{1.1.4}$$

This has the exact solution

$$x(t) = \frac{x_0}{[e^{-2t} + x_0^2(1 - e^{-2t})]^{1/2}}. \tag{1.1.5}$$

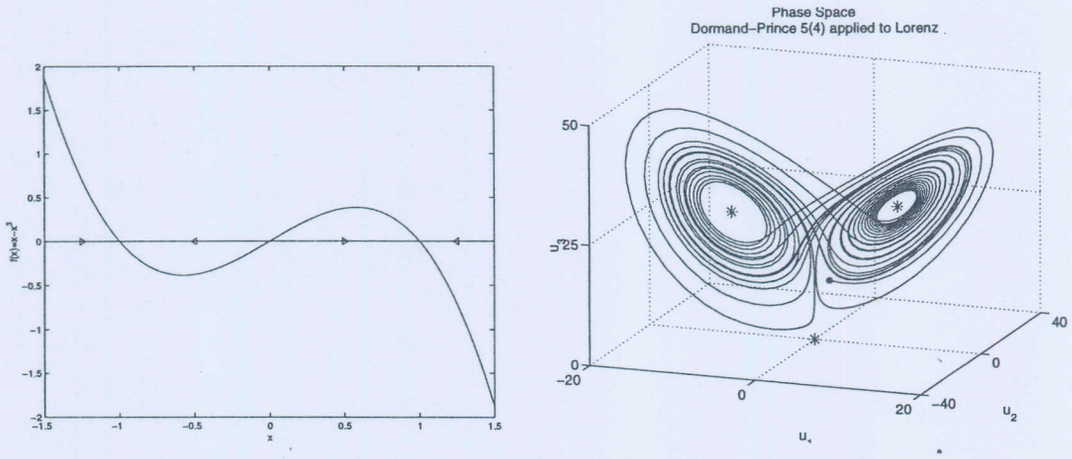


Figure 1: (i) Graph of $f(x) = x - x^3$ against x with the dynamics of (1.1.4) indicated on the x-axis. (ii) A numerically computed trajectory of the Lorenz equations (1.1.6).

From the graph of $f(x)$ (see Fig. 1), or the exact solution, it is easy to infer that for any $x_0 > 0$, $x(t) \to 1$ as $t \to \infty$, similarly $x(t) \to -1$ as $t \to \infty$ if $x_0 < 0$, and if $x_0 = 0$ then $x(t) = 0$ for all $t \in \mathbb{R}$.

(iii) The Lorenz equations in $\mathbb{R}^3$:

$$\begin{aligned} \dot{x} &= \sigma(y - x), \quad x(0) = x_0 \\ \dot{y} &= rx - y - xz, \quad y(0) = y_0 \\ \dot{z} &= xy - bz, \quad z(0) = z_0. \end{aligned} \tag{1.1.6}$$

These equations have no general closed form solution.

In Fig. 1 a numerical approximation to a trajectory over a very long time period is plotted. Since the classical error bound for the numerical approximation grows exponentially in time it is important to question the relationship of this figure to the true dynamics.

Lyapunov functions

It is often of importance to understand how functions of $x(t)$ change with time. By using the *generator* $\mathcal{L}$:

$$\mathcal{L}V = \sum_{i=1}^{d} f_i \frac{\partial V}{\partial x_i} = \langle f(x), \nabla V(x) \rangle, \tag{1.1.7}$$

we may achieve this. If $x(t)$ solves (1.1.1) and $V \in C^2(\mathbb{R}^d, \mathbb{R})$ then

$$\frac{d}{dt}\Big\{V(x(t))\Big\} = \mathcal{L}V(x(t)). \tag{1.1.8}$$

Note that $\mathcal{L}V(x(t)) = \langle \dot{x}, \nabla V(x) \rangle$, and often for deterministic dynamical systems $\frac{d}{dt}V(x(t))$ is calculated directly, without recourse to $\mathcal{L}$. We introduce $\mathcal{L}$ here in the deterministic case because of the central role of its generalisation in the stochastic case.

Also important is the adjoint operator $\mathcal{L}^*$, defined by $\langle f, \mathcal{L}g \rangle = \langle \mathcal{L}^* f, g \rangle$ with $\langle \cdot, \cdot \rangle$ the standard inner product on $\mathbb{R}^d$, so that in this case

$$\mathcal{L}^* V = -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}\Big(f_i V\Big).$$

This leads us to the Liouville equation for propagation of probability densities:

1.1.2 Lemma *If x_0 is a random variable with density $\rho_0(x)$ then $x(t)$ is a random variable which, under appropriate differentiability conditions on f and ρ_0, has density $\rho(x,t)$ which is a classical solution of the equation*

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= \mathcal{L}^* \rho, \quad (x,t) \in \mathbb{R}^d \times (0,T], \\ \rho(x,0) &= \rho_0(x), \quad x \in \mathbb{R}^d. \end{aligned}$$

Numerical methods

We consider numerical methods for the ODE (1.1.1) which, given $\Delta t > 0$, define approximations x_n to $x(t_n)$ at the equally spaced time intervals $t_n = n\Delta t \in [0,T]$.

Initially we introduce two numerical methods for the ODE (1.1.1). The first is the *forward Euler* method, sometimes called the *explicit Euler* method, and is defined as

$$x_{n+1} = x_n + \Delta t f(x_n). \tag{1.1.9}$$

There are many derivations of this method. From (1.1.2) we have

$$x(t_{n+1}) = x(t_n) + \int_{t_n}^{t_{n+1}} f(x(s))\, ds,$$

so letting $x_n = x(t_n)$ and approximating the integral

$$\int_{t_n}^{t_{n+1}} f(x(s))\, ds \approx \int_{t_n}^{t_{n+1}} f(x(t_n))\, ds = \Delta t f(x_n)$$

we obtain the forward Euler method.

The second method considered is the *backward Euler* method defined as

$$x_{n+1} = x_n + \Delta t f(x_{n+1}). \tag{1.1.10}$$

This method is sometimes called the *implicit Euler* method. Since x_{n+1} appears twice in (1.1.10) in general we have to solve a nonlinear equation to find x_{n+1} at each step. This can make the method much more computationally expensive to implement than the forward Euler method; nevertheless there are situations where the extra work is justified and we will illustrate some of these in these notes.

The backward Euler method can be derived in a similar fashion to the forward Euler method. Many other simple numerical methods such as the *trapezoidal rule* or *implicit midpoint* rule can also be derived by applying the corresponding numerical integration formula to the integral in (1.1.2). These methods are all examples of Runge-Kutta methods.

Also of interest are so-called *split step* or *partitioned methods*. Suppose $f(x) = f_1(x) + f_2(x)$, then we can define the *split-step Euler method* by

$$\left.\begin{aligned} x_* &= x_n + \Delta t f_1(x_n), \\ x_{n+1} &= x_* + \Delta t f_2(x_{n+1}). \end{aligned}\right\} \tag{1.1.11}$$

Such splitting methods are important in the solution of Hamiltonian systems and stochastic problems, and we will see examples of both.

1.1.3 Example (i) Applying the forward Euler method to the linear test problem (1.1.3) we obtain

$$x_{n+1} = x_n + \Delta t f(x_n) = x_n + \Delta t \lambda x_n = (1 + \Delta t \lambda) x_n.$$

Since this is true for arbitrary n we obtain the general solution

$$x_n = (1 + \Delta t \lambda)^n x_0.$$

(ii) Applying the forward Euler method to (1.1.4) we obtain

$$x_{n+1} = x_n + \Delta t x_n (1 - x_n^2).$$

In this case, as in general for nonlinear problems, there is no simple formula linking x_n directly to x_0.

(iii) Applying the backward Euler method to equation (1.1.4) we obtain

$$x_{n+1} = x_n + \Delta t x_{n+1} (1 - x_{n+1}^2).$$

Thus we need to solve a cubic equation to find x_{n+1}, given x_n.

(iv) Consider the separable Hamiltonian system

$$\dot{p} = \frac{\partial V}{\partial q}, \qquad \dot{q} = -\frac{\partial T}{\partial p},$$

where $p, q \in \mathbb{R}^d$ are generalised position and momenta, and $H(p,q) = T(p) + V(q)$ is preserved. We can apply the forward or backward Euler method directly to these problems. Alternatively write

$$\dot{x} = \begin{pmatrix} \dot{p} \\ \dot{q} \end{pmatrix} = \begin{pmatrix} \frac{\partial H}{\partial q} \\ -\frac{\partial H}{\partial p} \end{pmatrix} = \begin{pmatrix} \frac{\partial H}{\partial q} \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -\frac{\partial H}{\partial p} \end{pmatrix} := f_1(x) + f_2(x).$$

Now apply the split-step Euler method (1.1.11) which, after a little calculation, leads to

$$\left.\begin{aligned} p_{n+1} &= p_n + \Delta t \frac{\partial V}{\partial q}(q_n) \\ q_{n+1} &= q_n - \Delta t \frac{\partial T}{\partial p}(p_{n+1}). \end{aligned}\right\} \tag{1.1.12}$$

In this context the method is known as the *symplectic Euler method.* It essentially preserves the Hamiltonian structure of the dynamical system and has much better approximation properties for this type of problem than forward or backward Euler. Moreover, note that although the general formulation (1.1.11) is implicit, the symplectic Euler method (1.1.12) is explicit for this separable Hamiltonian problem [21].

In general when applying the backward Euler method, or any other implicit method, there are nonlinear equations to solve at each step. We will briefly consider how these equations can be solved in practice by considering the backward Euler method in one dimension:

$$x_{n+1} = x_n + \Delta t f(x_{n+1}), \quad x_n \in \mathbb{R}. \tag{1.1.13}$$

We could try to solve this by the simple iteration

$$x_{n+1}^{k+1} = x_n + \Delta t f(x_{n+1}^k), \quad x_{n+1}^0 = x_n,$$

and let $x_{n+1} = \lim_{k\to\infty} x_{n+1}^k$ (if the limit exists). Of course, in practice we cannot iterate to infinity so we stop iterating if $|x_{n+1}^k - x_{n+1}^{k-1}| < \tau$ for some small tolerance $\tau > 0$. However, this iteration has the drawback that $\lim_{k\to\infty} x_{n+1}^k$ need not exist if Δt is too large. Indeed ensuring that this iteration is convergent often imposes quite seyere restrictions on the step-size Δt.

Thus it is often preferable to use the more sophisticated Newton iteration. For $g \in C(\mathbb{R}, \mathbb{R})$, Newton's method for solving $g(x) = 0$ is given by iterating for $k = 0, 1, 2, \ldots$,

$$x^{k+1} = x^k - \frac{g(x^k)}{g'(x_k)}.$$

So to solve (1.1.13) we let

$$g(x) = x - x_n - \Delta t f(x),$$

so that finding x_{n+1} is equivalent to solving $g(x) = 0$ which by Newton's method we can do by iterating

$$x_{n+1}^{k+1} = x_{n+1}^k - \frac{x_{n+1}^k - x_n - \Delta t f(x_{n+1}^k)}{1 - \Delta t f'(x_{n+1}^k)}.$$

Thus when applying an implicit method we have to perform an inner iteration in k at each step of the outer iteration in n.

One way to avoid this is to set $x_{n+1}^0 = x_n$ and take $k = 1$, to obtain the *linearised Euler method*

$$x_{n+1} = x_n + \Delta t \frac{f(x_n)}{1 - \Delta t f'(x_n)}.$$

These ideas generalise to arbitrary dimension.

Bibliographical remark Throughout the following we minimise the size of the bibliography by often referring to [24]. Although it is often not the original source of the material we reference, it has a comprehensive bibliography and pointers to the relevant literature at the end of every chapter.

1.2 Dynamical systems

There are several definitions of "dynamical system" in use. In the context of the ODE (1.1.1), many definitions require both a unique solution for all $t \in \mathbb{R}$ and all initial conditions $x_0 \in \mathbb{R}^d$, and also that for any fixed $t \in \mathbb{R}$ the solution is continuous with respect to the initial data x_0. However, many interesting problems such as (1.1.4) and the Lorenz equations, which possess bounded solutions in the limit as $t \to \infty$, have solutions which become unbounded in finite negative time, and so by such definitions would not be dynamical systems.

1.2.1 Example Consider (1.1.4) with $|x_0| > 1$. Then using the exact solution (1.1.5) we see that $|x(t)| \to \infty$ as $t \to \frac{1}{2}\ln(1 - \frac{1}{x_0^2}) < 0$.

Thus this definition is too restrictive, and we will work with a weaker concept. The differential equation (1.1.1) is said to define a *semi-dynamical system* if it has a unique solution for all $t \geqslant 0$ and all initial conditions $x_0 \in \mathbb{R}^d$. We will work with semi-dynamical systems throughout, but for brevity will omit the "semi" and refer to them as dynamical systems. Note that we make no reference to continuity with respect to initial data, though for (1.1.1) with continuous f uniqueness of solutions implies continuity with respect to initial data [7].

Dynamical systems defined by ODEs are characterised by having a continuous time variable. We can also have dynamical systems with a discrete time variable. The continuous map $g : \mathbb{R}^d \to \mathbb{R}^d$ defines a discrete dynamical system via the iteration

$$x_{n+1} = g(x_n). \tag{1.2.1}$$

Since numerical methods are iterated in this way they are good candidates to be considered as discrete dynamical systems.

We will be interested in both dynamical systems defined by ODEs and in discrete dynamical systems. Our discrete dynamical systems will arise from numerical solutions of ODEs. We will analyse the behaviour of numerical methods by using dynamical systems techniques to compare the dynamical systems defined by an ODE and its numerical solution.

Evolution operator

To be able to analyse both continuous and discrete time dynamical systems together, we now set up general notation for a dynamical system evolving in a state space $\mathbb{X}$, which will usually be $\mathbb{R}^d$, and time space $\mathbb{T}$, where usually $\mathbb{T} = \mathbb{R}^+$ for the continuous time case, and $\mathbb{T} = \mathbb{Z}^+$ in the discrete time case; the choices $\mathbb{T} = \mathbb{R}$ and $\mathbb{T} = \mathbb{Z}$ will also occur occasionally. Our primary motivation here is the study of (1.1.1) and its discrete time counterparts (1.1.9) and (1.1.10).

1.2.2 Definition A *dynamical system* on a space $\mathbb{X}$ with time $\mathbb{T}$ is a mapping

$$\varphi : \mathbb{T} \times \mathbb{X} \to \mathbb{X}, \quad (t, x_0) \mapsto \varphi(t, x_0), \tag{1.2.2}$$

where the mapping φ satisfies the *semi-group property* that $\varphi(0, x) = x$ for all $x \in \mathbb{X}$ and

$$\varphi(t + s, x_0) = \varphi(t, \varphi(s, x_0)) \quad \forall s, t \in \mathbb{T},\ x \in \mathbb{X}.$$

It will sometimes be helpful to write

$$\varphi(t,x) = \varphi^t x,$$

and we write φ for φ^1. In this notation the semi-group property reads $\varphi^0 = \mathrm{id}_{\mathbb{X}}$ and

$$\varphi^{t+s} = \varphi^t \circ \varphi^s.$$

For a set $B \subset \mathbb{X}$ we write $\varphi(t,B)$ for

$$\varphi(t,B) = \bigcup_{x\in B} \varphi(t,x).$$

The mapping φ is known as the *evolution operator*. For the ODE (1.1.1) we have $\varphi(t,x_0) = \varphi^t x_0 = x(t)$, the solution at time t with initial condition $x(0) = x_0$. For the map (1.2.1) we have $\varphi(n,x_0) = \varphi^n x_0 = x_n$, the solution after n iterates of the map with initial condition x_0.

Examples

1.2.3 Example For (1.1.3) we have $\mathbb{X} = \mathbb{R}$ and $\mathbb{T} = \mathbb{R}$ with

$$\varphi(t,x_0) = \varphi^t x_0 = e^{\lambda t} x_0 = x(t).$$

1.2.4 Example From Example 1.1.3(i), applying the forward Euler method to (1.1.3), we have $\mathbb{X} = \mathbb{R}$ and $\mathbb{T} = \mathbb{Z}$ with

$$\varphi(n,x_0) = \varphi^n x_0 = (1+\Delta t\lambda)^n x_0 = x_n.$$

1.2.5 Example For (1.1.4) we have $\mathbb{X} = \mathbb{R}$ and $\mathbb{T} = \mathbb{R}^+$ with

$$\varphi(t,x_0) = \varphi^t x_0 = \frac{x_0}{[e^{-2t} + x_0^2(1-e^{-2t})]^{1/2}} = x(t).$$

1.2.6 Example For the forward Euler method applied to (1.1.4) we have $\mathbb{X} = \mathbb{R}$ and $\mathbb{T} = \mathbb{Z}^+$ with

$$\varphi(1,x_n) = \varphi^1 x_n = \varphi x_n = x_n + \Delta t f(x_n) = x_n + \Delta t(x_n - x_n^3) = x_{n+1}.$$

1.2.7 Example For the backward Euler method applied to (1.1.4) we have $\mathbb{X} = \mathbb{R}$ and $\mathbb{T} = \mathbb{Z}^+$ with

$$\varphi(1,x_n) = \varphi^1 x_n = \varphi x_n = x_n + \Delta t f(\varphi x_n) = x_n + \Delta t(\varphi x_n - (\varphi x_n)^3) = x_{n+1}.$$

It is not immediately apparent whether φ is well-defined in this case. We discuss this issue below.

Existence and uniqueness

In defining dynamical systems via the mapping (1.2.2) we are implicitly assuming existence and uniqueness of solutions. In practice, existence and uniqueness of solutions will need to be verified in different scenarios.

1.2.8 Example Consider the ODE

$$\dot{x} = x^2, \quad x(0) = 1,$$

which has exact solution

$$x(t) = \frac{1}{1/x_0 - t}.$$

Thus $x(t) \to \infty$ as $t \to 1/x_0$ for $x_0 > 0$, so this differential equation does not define a dynamical system.

1.2.9 Definition A function $f : \mathbb{X} \to \mathbb{X}$ is *Lipschitz* on $B \subset \mathbb{X}$ with Lipschitz constant L_B if

$$\|f(x) - f(y)\| \leqslant L_B \|x - y\|, \quad \forall x, y \in B.$$

If f is Lipschitz on $\mathbb{X}$ then f is *globally Lipschitz*. If f is Lipschitz on every bounded subset of $\mathbb{X}$, then f is *locally Lipschitz*.

Clearly globally Lipschitz implies locally Lipschitz. Note also that if f is differentiable on $\mathbb{X}$, then f must be locally Lipschitz. The following lemmas will be useful.

1.2.10 Lemma (Continuous Gronwall Lemma) *Let $z(t)$ satisfy*

$$\dot{z} \leqslant az + b, \quad z(0) = z_0,$$

for constants $a \neq 0$ and $b \in \mathbb{R}$. Then

$$z(t) \leqslant e^{at} z_0 + \frac{b}{a}(e^{at} - 1).$$

Proof Integrate the differential inequality, see for example [24]. □

1.2.11 Lemma (Discrete Gronwall Lemma) *Let $C \in (0, 1)$ and $D_n \in \mathbb{R}$ and suppose $Z_n \in \mathbb{R}$ satisfies*

$$Z_{n+1} \leqslant CZ_n + D_n, \quad \forall n \geqslant 0.$$

Then

$$Z_n \leqslant Z_0 C^n + \sum_{j=1}^{n} C^{n-j} D_{j-1}.$$

If $D_n \equiv D$ is constant then

$$Z_n \leqslant \frac{D}{1 - C}(1 - C^n) + Z_0 C^n. \tag{1.2.3}$$

Proof By induction, see [24] for example. □

1.2.12 Theorem *If f is globally Lipschitz on $\mathbb{R}^d$ then the ODE* (1.1.1) *has a unique solution for all $t \in \mathbb{R}$ and any $x_0 \in \mathbb{R}^d$. Thus φ^t is well defined for all $t \in \mathbb{R}$, and* (1.1.1) *defines a dynamical system on $\mathbb{X} = \mathbb{R}^d$, with $\mathbb{T} = \mathbb{R}$ or $\mathbb{R}^+$. Moreover,*

$$e^{-Lt}\|x_0 - y_0\| \leqslant \|\varphi^t x_0 - \varphi^t y_0\| \leqslant e^{Lt}\|x_0 - y_0\|. \tag{1.2.4}$$

Proof Uses Picard iteration, see for example [24]. The final bound follows from

$$\frac{1}{2}\frac{d}{dt}\|x(t) - y(t)\|^2 = \langle f(x) - f(y), x - y\rangle,$$

and

$$-L\|x - y\|^2 \leqslant \langle f(x) - f(y), x - y\rangle \leqslant L\|x - y\|^2,$$

using the Continuous Gronwall Lemma 1.2.10. □

Note that (1.2.4) bounds the rate of both convergence and divergence of trajectories. In particular it means that trajectories cannot cross or merge as, if $x_0 \neq y_0$, then $x(t) \neq y(t)$ for all $t \in \mathbb{R}$.

Unfortunately this theorem is of limited application as few f encountered in practice are globally Lipschitz, though many are locally Lipschitz. For example, both (1.1.4) and the Lorenz equations (1.1.6) have differentiable and hence locally Lipschitz f, but f is not globally Lipschitz.

However, if f is locally Lipschitz it can be shown that solutions are unique whilst they exist, and a solution $x(t)$ can only fail to exist if it blows up in finite time, i.e. $\|x(t)\| \to \infty$ as $t \to T < \infty$. Often it is possible to prevent such blow up. The generator (1.1.7) will be useful for this purpose.

1.2.13 Assumption (Generator Assumption) There is a function $V \in C(\mathbb{X}, \mathbb{R})$ with $\lim_{\|x\|\to\infty} V(x) = \infty$, and real numbers $a, b, d \in (0, \infty)$, $c \in \mathbb{R}$ such that $V(0) < d/a$ and

$$\mathcal{L}\{V(x)\} \leqslant b\|x\|^c[d - aV(x)], \tag{1.2.5}$$

where $\mathcal{L}$ is the generator for (1.1.1) given by (1.1.7).

1.2.14 Theorem *If f is locally Lipschitz and satisfies Assumption* 1.2.13 *then the ODE* (1.1.1) *has a unique solution for all $t \geqslant 0$, and any $x_0 \in \mathbb{R}^d$. Thus φ^t is well defined for all $t \in \mathbb{R}^+$, and* (1.1.1) *defines a dynamical system on $\mathbb{X} = \mathbb{R}^d$ with $\mathbb{T} = \mathbb{R}^+$.*

Proof

$$\frac{d}{dt}\{V(x(t)\} = \mathcal{L}V(x(t)) \leqslant b\|x(t)\|^c[d - aV(x(t))].$$

Hence $\dot{V} \leqslant 0$ when $V(x(t)) \geqslant d/a$. It follows that

$$\limsup_{t\to\infty} V(x(t)) \leqslant \max\left\{V(x(0)), \frac{d}{a}\right\} := M.$$

But since $\lim_{\|x\|\to\infty} V(x) = \infty$ there exists $R \geqslant 0$ such that $V(x) > M$ for all x such that $\|x\| > R$. Thus $\limsup_{t\to\infty} \|x(t)\| \leqslant R$, and solutions cannot blow up. The rest of the proof mimics the globally Lipschitz case. □

We will also consider f under the following assumption.

1.2.15 Assumption (One-Sided Lipschitz Condition) There exists $c > 0$ such that

$$\langle f(u) - f(v), u - v \rangle \leqslant c\|u - v\|^2 \tag{1.2.6}$$

for all $u, v \in \mathbb{R}^d$.

1.2.16 Example Let $f(x) = x - x^3$ as in (1.1.4). Then f is differentiable and hence locally Lipschitz, but is not globally Lipschitz. Now

$$\begin{aligned}\langle f(u) - f(v), u - v \rangle &= (u - v)(u - u^3 - v + v^3) \\ &= (u - v)^2 \left(1 - \frac{1}{2}u^2 - \frac{1}{2}(u + v)^2 - \frac{1}{2}v^2\right) \\ &\leqslant (u - v)^2,\end{aligned}$$

so this function does satisfy a one-sided Lipschitz condition.

1.2.17 Theorem *If f is locally Lipschitz and satisfies Assumption* 1.2.15, *then* (1.1.1) *defines a dynamical system on $\mathbb{X} = \mathbb{R}^d$ with $\mathbb{T} = \mathbb{R}^+$, and φ^t satisfies*

$$\|\varphi^t x_0 - \varphi^t y_0\| \leqslant e^{ct}\|x_0 - y_0\|.$$

Proof Note that the final bound follows from

$$\frac{1}{2}\frac{d}{dt}\|x(t) - y(t)\|^2 = \langle f(x) - f(y), x - y \rangle \leqslant c\|x - y\|^2,$$

using the Continuous Gronwall Lemma 1.2.10. □

The situation for maps defined by explicit numerical methods is quite different. Given $f : \mathbb{R}^d \to \mathbb{R}^d$ the forward Euler method has $\varphi x = x + \Delta t f(x)$ for any $x \in \mathbb{R}^d$ and hence the map can be iterated infinitely many times, so defines a dynamical system.

However, for implicit numerical methods such as the backward Euler method we need to verify existence and uniqueness of solutions.

1.2.18 Theorem *Suppose f is globally Lipschitz with Lipschitz constant L and $L\Delta t < 1$, then the backward Euler method applied to* (1.1.1) *defines a dynamical system on $\mathbb{X} = \mathbb{R}^d$ with $\mathbb{T} = \mathbb{Z}$ or $\mathbb{Z}^+$.*

Proof Let $g(x) = x_n + \Delta t f(x)$. Then $g(x_{n+1}) = x_{n+1}$, and so we are required to show that for any $x_n \in \mathbb{R}^d$ the function $g(x)$ has a unique fixed point. Now

$$\|g(x) - g(y)\| = \Delta t\|f(x) - f(y)\| \leqslant L\Delta t\|x - y\|,$$

and since $L\Delta t < 1$, g is a contraction mapping of $\mathbb{R}^d$ into itself, and so has a unique fixed point. Backward in time existence and uniqueness follow because x_n is given explicitly from x_{n+1}. □

As already noted, not many f are globally Lipschitz. Hence the following is important.

1.2.19 Lemma *Suppose f satisfies a one-sided Lipschitz condition* (1.2.6) *and $c\Delta t < 1$, then the equation*

$$a - \Delta t f(a) = b, \tag{1.2.7}$$

has a unique solution $a \in \mathbb{R}^d$ for any $b \in \mathbb{R}^d$.

Proof Let $g(x) = x - b - \Delta t f(x)$. Then

$$\langle g(x) - g(y), x - y\rangle = \langle x - y - \Delta t(f(x) - f(y)), x - y\rangle \geqslant (1 - c\Delta t)\|x - y\|^2.$$

Now since $1 - c\Delta t > 0$ we can apply the uniform monotonicity theorem to determine the existence of a unique solution. See [24] for details. □

1.2.20 Corollary *Suppose f satisfies a one-sided Lipschitz condition* (1.2.6) *and $c\Delta t < 1$, then the backward Euler method applied to* (1.1.1) *defines a dynamical system with $\mathbb{X} = \mathbb{R}^d$, $\mathbb{T} = \mathbb{Z}^+$.*

1.2.21 Example Corollary 1.2.20 combined with Example 1.2.16 implies that the evolution operator φ defined in Example 1.2.7 for the backward Euler method applied to (1.1.4) is well-defined on $\mathbb{R}$ provided $\Delta t < 1$.

In some situations it is only possible to prove existence, and not uniqueness, of a numerical solution. See Lemma 1.4.13 for example.

1.3 Invariant sets

Consider two dynamical systems defined by mappings φ, $\varphi_{\Delta t}$ (where φ is the evolution operator for an ODE of the form (1.1.1) and $\varphi_{\Delta t}$ is the evolution operator for its numerical approximation). To compare these dynamical systems we compare their dynamic properties. (This contrasts with the traditional numerical analysis approach of comparing errors between trajectories: $\|\varphi^n x_0 - \varphi^n_{\Delta t} x_0\|$.) The simplest features of dynamical systems to compare are invariant sets, and the simplest invariant sets are fixed points.

1.3.1 Definition A point $\bar{u} \in \mathbb{X}$ is a *fixed point* of $\varphi : \mathbb{T} \times \mathbb{X} \to \mathbb{X}$ if $\varphi^t \bar{u} = \varphi(t, \bar{u}) = \bar{u}$ for all $t \in \mathbb{T}$.

1.3.2 Example (i) Consider (1.1.3), with $\lambda \neq 0$. Then it follows from Example 1.2.3 that $\bar{u} = 0$ is the unique fixed point.

(ii) Consider the forward Euler method applied (1.1.3), with $\lambda \neq 0$. Again $\bar{u} = 0$ is the unique fixed point.

It is clear from (1.1.1) that $\bar{u} \in \mathbb{R}^d$ is a fixed point of the differential equation if and only if $f(\bar{u}) = 0$. Also from (1.1.9) and (1.1.10) we see that $\bar{u}$ is a fixed point for the forward Euler or backward Euler methods if and only if $f(\bar{u}) = 0$. So in the case of numerical approximation by forward or backward Euler the fixed points of φ and $\varphi_{\Delta t}$ agree. This is not true for all numerical methods: many methods admit so-called *spurious fixed points* $\bar{u}_{\Delta t}$ with $\varphi^n_{\Delta t} \bar{u}_{\Delta t} = \bar{u}_{\Delta t}$ for all $n \in \mathbb{Z}$ but $f(\bar{u}_{\Delta t}) \neq 0$.

We now consider more general invariant sets.

1.3.3 Definition A bounded set $B \subset \mathbb{X}$ is *positively invariant* if $\varphi^t B \subset B$ for all $t \in \mathbb{T}$, such that $t \geqslant 0$ and *invariant* if also $B \subset \varphi^t B$ for all $t \in \mathbb{T}$ such that $t \geqslant 0$.

Note that a set B is thus invariant if $\varphi^t B = B$ for all $t \in \mathbb{T}$. If B is an invariant set then for any $u_0 \in B$ there exists $u : \mathbb{T} \to \mathbb{X}$ such that for any $0 < t \in \mathbb{T}$ and $0 > s \in \mathbb{T}$, $u(t) = \varphi^t u_0$, $\varphi^{-s} u(s) = u_0$ and $\varphi^t u(s) = u(s+t)$. The set $\{u(t) : t \in \mathbb{T}\}$ is called a *complete orbit* of u_0. Indeed the complete orbit of any point $u_0 \in \mathbb{X}$ is itself an invariant set, and a set $B \in \mathbb{X}$ is invariant if and only if every point $x \in B$ is on a complete orbit which is contained in B; see for example [24].

1.3.4 Example Consider (1.1.4). The fixed points are $\{-1\}$, $\{0\}$, $\{1\}$. Additionally $(-1, 0)$ and $(0, 1)$ are both invariant sets, as is the union of any combination of these five invariant sets.

1.3.5 Example Consider

$$\left.\begin{aligned} \dot{r} &= r(1-r^2), & r &\geqslant 0, \\ \dot{\theta} &= 1, & \theta &\in [0, 2\pi]. \end{aligned}\right\} \tag{1.3.1}$$

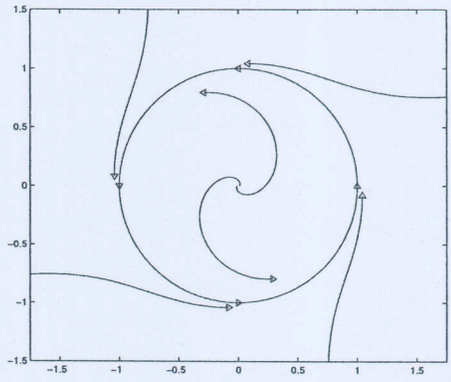


Figure 2: Trajectories of (1.3.1).

Here the bounded invariant sets include $\{(r, \theta) : r = 1\}$, $\{(r, \theta) : 0 < r < 1\}$, $\{(r, \theta) : r < 1\}$, and the complete orbit through any point on the unit disc. There are thus uncountably many bounded invariant sets in this case.

The set $\{(r, \theta) : r = 1\}$ in the previous example is a *periodic orbit.* Other invariant sets include invariant tori and strange attractors. Often invariant sets which are "stable" are of more interest than unstable invariant sets and this leads us to the concept of attractors.

1.4 Attractors

An attractor is defined as an invariant set $\mathcal{A}$ such that $\varphi^t B$ approaches $\mathcal{A}$ as $t \to \infty$ for all bounded $B \subset \mathbb{X}$.

1.4.1 Definition For sets $A, B \subseteq \mathbb{X}$ the Hausdorff semi-distance $\mathrm{dist}(A, B)$ is defined by

$$\mathrm{dist}(A, B) = \sup_{a \in A} \inf_{b \in B} \|a - b\|.$$

Note that Definition 1.4.1 only gives a semi-distance: $\mathrm{dist}(A, B)$ will be small if every point $a \in A$ is close to some point $b \in B$; but there may be points $b \in B$ which are not close to A. In particular, if $A \subset B$ then $\mathrm{dist}(A, B) = 0$ but $\mathrm{dist}(B, A)$ may be large.

1.4.2 Example Let $A = \{0\}$, and $B = [0, 1]$. Then $\mathrm{dist}(A, B) = 0$, but $\mathrm{dist}(B, A) = 1$.

1.4.3 Definition The set $\mathcal{A} \subset \mathbb{X}$ is a *global attractor* if,

- $\mathcal{A}$ is a compact set;
- $\varphi^t \mathcal{A} = \mathcal{A}$ for all $t \geqslant 0$ (invariance);
- for all bounded B

$$\lim_{t \to \infty} \mathrm{dist}(\varphi^t B, \mathcal{A}) = 0.$$

1.4.4 Definition K is an *absorbing set* if, for all bounded $B \subset \mathbb{X}$, there exists $t^* = t^*(B) \geqslant 0$:

$$\varphi^t B \subseteq K \quad \forall t \geqslant t^*.$$

The following theorem (see [6, 28, 24]) is valid in both discrete and continuous time, with $t, T \in \mathbb{T}$.

1.4.5 Theorem *If a dynamical system has a compact absorbing set K, then*

$$\mathcal{A} = \bigcap_{t \geqslant 0} \varphi^t K$$

is a global attractor.

Note that in this case

$$\mathcal{A} = \bigcap_{t \geqslant 0} \varphi^t K = \bigcap_{t \geqslant 0} \overline{\bigcup_{T \geqslant t} \varphi^T K} := \Lambda(K),$$

where $\Lambda(K)$ is the *ω-limit set* of K. Thus the global attractor is the ω-limit set of the absorbing set. Note that the form of the ω-limit set $\Lambda(K)$ simplifies, because the absorbing set maps the dynamics into itself.

Finding invariant sets directly is usually very hard, especially in the chaotic case. However, Theorem 1.4.5 allows us to infer the existence of an attractor whenever we have an absorbing set.

1.4.6 Example For Example 1.2.3 with $\lambda < 0$ we have $\varphi^t B = e^{\lambda t} B$, and it is easy to verify that any interval $B = [-a, a]$ is an absorbing set. Hence

$$\mathcal{A} = \Lambda(B) = \bigcap_{t \geqslant 0} \varphi^t B = \bigcap_{t \geqslant 0} e^{\lambda t}[-a, a] = \{0\},$$

and thus the global attractor is the (unique) fixed point.

1.4.7 Example For Example 1.2.4 with $\lambda < 0$ and $|1+\Delta t\lambda| < 1$ we have $\varphi^n B = (1+\Delta t\lambda)^n B$, and again any interval $B = [-a, a]$ is an absorbing set. Hence

$$\mathcal{A} = \Lambda(B) = \bigcap_{n\geqslant 0} \varphi^n B = \bigcap_{n\geqslant 0} (1+\Delta t\lambda)^n [-a, a] = \{0\},$$

and thus the global attractor is the (unique) fixed point. Notice that the condition for the numerical solution to have the same attractor as in the previous example is that $\Delta t < -2/\lambda$.

1.4.8 Example For Example 1.2.5 with $B = [a, b]$ we have

$$\varphi^t B = \left[\frac{a}{[e^{-2t} + a^2(1-e^{-2t})]^{1/2}}, \frac{b}{[e^{-2t} + b^2(1-e^{-2t})]^{1/2}}\right],$$

and provided $a < -1$ and $b > 1$, $B = [a, b]$ is an absorbing set. Hence

$$\begin{aligned}\mathcal{A} = \Lambda(B) &= \bigcap_{t\geqslant 0} \varphi^t B = \bigcap_{t\geqslant 0} \left[\frac{a}{[e^{-2t} + a^2(1-e^{-2t})]^{1/2}}, \frac{b}{[e^{-2t} + b^2(1-e^{-2t})]^{1/2}}\right] \\ &= [-1, 1],\end{aligned}$$

and thus the global attractor is the interval $[-1, 1]$.

1.4.9 Example For Example 1.3.5 the global attractor is the disc $\{r \leqslant 1\}$.

Now we study the ODE (1.1.1), together with the backward Euler approximation of it. Recall the Generator Assumption 1.2.13.

1.4.10 Theorem *Let f be locally Lipschitz and satisfy Assumption* 1.2.13. *Then the ODE* (1.1.1) *defines a dynamical system on $\mathbb{R}^d$ and has a global attractor.*

Proof By Theorem 1.2.14 the ODE (1.1.1) defines a dynamical system on $\mathbb{R}^d$, and it remains to prove that it has an absorbing set to establish that it has a global attractor.

Since $V(0) < d/a$, there exists $R > 0$ such that $V(x) < d/a$ for all x such that $\|x\| < R$. Now let $Y(t) = V(x(t))$ and suppose $Y(t) \geqslant d/a$. Then

$$\dot{Y}(t) \leqslant b\|x\|^c[d - aY(t)] \leqslant bR^c[d - aY(t)].$$

This implies that $B_\varepsilon = \{x \in \mathbb{R}^d : V(x) \leqslant d/a + \varepsilon\}$ is forward invariant for any $\varepsilon > 0$. To establish the existence of an absorbing set, and hence a global attractor, it remains to show that trajectories enter such a set in finite time.

Let B be a bounded set and let $x_0 \in B$ with $V(x_0) > d/a$. Then, whilst $x(t) \notin B_\varepsilon$,

$$\dot{Y} + abR^cY(t) \leqslant bdR^c,$$

and integrating from $[0, t]$

$$Y(t) \leqslant \frac{d}{a} + e^{-abR^c t}\left[Y(0) - \frac{d}{a}\right].$$

A simple calculation shows that if

$$t > \frac{1}{abR^c} \ln\left(\frac{1}{\varepsilon}\left[Y(0) - \frac{d}{a}\right]\right),$$

then $Y(t) < d/a + \varepsilon$, hence

$$\left\{x : V(x) \leqslant \frac{d}{a} + \varepsilon\right\}$$

is an absorbing set for any $\varepsilon > 0$. Finally note that the attractor $\mathcal{A} \subset \{x : V(x) \leqslant d/a\}$ (since if not we get a contradiction by choosing ε sufficiently small.) □

Note that $Y(t)$ enters B_ε exponentially fast. The random analogue of this fact will give exponential convergence to invariant measures. See Section 2.3.

We will also employ the following assumption which implies Assumption 1.2.13.

1.4.11 Assumption (Dissipativity Condition) There exists α, $\beta > 0$ and $p \geqslant 1$ such that

$$\langle x, f(x)\rangle \leqslant \alpha - \beta\|x\|^{2p}, \quad \forall x \in \mathbb{R}^d.$$

1.4.12 Theorem *Let f be locally Lipschitz and satisfy Assumption* 1.4.11. *Then the ODE* (1.1.1) *defines a dynamical system on $\mathbb{R}^d$ and has a global attractor contained in $\{x : \|x\|^{2p} \leqslant \alpha/\beta\}$.*

Proof Let $V(x) = \|x\|^{2p} = (\sum_{i=1}^d x_i^2)^p$. Then $\partial V/\partial x_i = 2p\|x\|^{2p-2}x_i$, and

$$\mathcal{L}V(x) = \langle f(x), \nabla V(x)\rangle = 2p\|x\|^{2p-2}\langle f(x), x\rangle \leqslant 2p\|x\|^{2p-2}[\alpha - \beta V(x)].$$

Thus Assumption 1.2.13 holds and the result follows by Theorem 1.4.10. □

We now move on to consider the backward Euler method (1.1.10). One complication we will encounter here is that, whilst a solution is guaranteed to exist for all time-steps positive, it is no longer unique. To be precise we need to extend the concept of dynamical system to allow set-valued sequences, which we refer to as *generalised dynamical systems* [24], generalizing Definition 1.2.2 in a fairly straightforward way. The following lemma underpins this existence theory.

1.4.13 Lemma *Let Assumption* 1.4.11 *hold, and let $b \in \mathbb{R}^d$ be given. Then the equation*

$$a - \Delta t f(a) = b,$$

has at least one solution $a \in \mathbb{R}^d$.

Proof Let $g(x) = x - b - \Delta t f(x)$. Then

$$\begin{aligned}\langle g(x), x\rangle &= \|x\|^2 - \langle b, x\rangle - \Delta t\langle f(x), x\rangle,\\ &\geqslant \Delta t\beta\|x\|^{2p} + \|x\|^2 - \|b\|\|x\| - \alpha\Delta t.\end{aligned}$$

Thus for $R > 0$ sufficiently large $\langle g(x), x\rangle > 0$ for all x such that $\|x\| = R$ and the result follows on applying the Brouwer fixed-point theorem. See [24] for details. □

The definitions of global attractors and related concepts for generalised dynamical systems are analogous to those for dynamical systems and similar properties hold [24].

1.4.14 Theorem *Let Assumption* 1.4.11 *hold, then the backward Euler method* (1.1.10) *defines a generalised dynamical system on* $\mathbb{R}^d$ *and has a global attractor* $\mathcal{A}_{\Delta t}$.

Proof The existence of a generalised dynamical system follows from Lemma 1.4.13. To deduce the existence of an attractor, consider

$$x_{n+1} - \Delta t f(x_{n+1}) = x_n$$

hence

$$\begin{aligned} \|x_{n+1}\|^2 &= \|x_n\|^2 + 2\Delta t \langle x_{n+1}, f(x_{n+1})\rangle - \Delta t^2 \|f(x_{n+1})\|^2 \\ &\leqslant \|x_n\|^2 + 2\Delta t(\alpha - \beta \|x_{n+1}\|^{2p}). \end{aligned}$$

Now given any $\varepsilon > 0$, assume there exists $x_0 \in \mathbb{R}^d$ such that $\|x_n\|^{2p} > \alpha/\beta + \varepsilon$ for all $n \geqslant 0$. Then

$$\|x_{n+1}\|^2 - \|x_n\|^2 \leqslant -2\Delta t \beta \varepsilon,$$

hence

$$\|x_n\|^2 \leqslant \|x_0\|^2 - 2n\Delta t \beta \varepsilon,$$

which supplies a contraction for n sufficiently large. Thus

$$\{x : \|x\|^{2p} \leqslant \alpha/\beta + \varepsilon\},$$

is an absorbing set for any $\varepsilon > 0$ and there is a global attractor $\mathcal{A}_{\Delta t} \subset \{x : \|x\|^{2p} \leqslant \alpha/\beta\}$. □

An important point to note in the previous theorem is that the backward Euler method has preserved the Lyapunov function of the continuous problem. More can be said about the attractor from Theorem 1.4.14:

1.4.15 Theorem *Let Assumptions* 1.4.11 *hold and let* $\mathcal{A}$ *be the global attractor of the dynamical system defined by* (1.1.1) *and* $\mathcal{A}_{\Delta t}$ *the global attractor for the backward Euler method applied to* (1.1.1) *then*

$$\lim_{\Delta t \to 0} \operatorname{dist}(\mathcal{A}_{\Delta t}, \mathcal{A}) = 0.$$

Proof See for example [24]. □

Recall that in general the backward Euler method only defines a generalised dynamical system for this problem. However, for Δt sufficiently small we can show that it defines a dynamical system on a neighbourhood of the attractor.

1.4.16 Theorem *Let Assumption* 1.4.11 *hold and let* L *be the Lipschitz constant for* f *on* $K = \{x : \|x\|^{2p} \leqslant \alpha/\beta\}$. *If* $L\Delta t < 1$ *then the backward Euler method applied to* (1.1.1) *defines a dynamical system on* K, *with* $\mathbb{T} = \mathbb{Z}^+$, *and for* x_0, $y_0 \in K$

$$(1 - L\Delta t)^n \|x_0 - y_0\| \leqslant \|x_n - y_n\| \leqslant (1 + L\Delta t)^n \|x_0 - y_0\|. \tag{1.4.1}$$

The inequalities (1.4.1) are a discrete analogue of (1.2.4) and bound the rate of convergence/divergence of trajectories, and prevent trajectories from merging or crossing in K.

Note that other conditions, for example the one-sided Lipschitz condition, can be used to infer that the backward Euler method defines a dynamical system on $\mathbb{R}^d$.

1.4.17 Example Consider the backward Euler method applied to (1.1.4). We showed in Example 1.2.21 that for $\Delta t < 1$, the map φ is well-defined, so that this defines a dynamical system on $\mathbb{R}$. Note that

$$\langle f(x), x\rangle = x^2 - x^4 = \frac{1}{2} - \frac{1}{2}(1-x^2)^2 - \frac{1}{2}x^4 \leqslant \frac{1}{2}(1-x^4)$$

and so Assumption 1.4.11 is satisfied, and hence Theorem 1.4.14 implies the existence of an attractor $\mathcal{A}_{\Delta t}$. Now Theorem 1.4.15 implies that

$$\lim_{\Delta t \to 0} \operatorname{dist}(\mathcal{A}_{\Delta t}, \mathcal{A}) = 0.$$

as $\Delta t \to 0$. In fact, a little algebra shows that $\mathcal{A}_{\Delta t} = \mathcal{A} = [-1, 1]$ for any $\Delta t < 1$.

The behaviour of the forward Euler method applied to the same ODE is very different as the following example shows.

1.4.18 Example Consider the forward Euler method applied to (1.1.4). Then

$$x_{n+1} = x_n + \Delta t(x_n - x_n^3) = [1 + \Delta t - \Delta t x_n^2]x_n.$$

Thus if $x_n^2 > 1 + 2/\Delta t$, then $[1 + \Delta t - \Delta t x_n^2] < -1$ and $|x_{n+1}| > |x_n|$. It follows that given any $\Delta t > 0$ for all x_0 with

$$x_0^2 > 1 + \frac{2}{\Delta t},$$

then $\lim_{n\to\infty} \|x_n\| = \infty$. Thus in this case there is no absorbing set and hence no global attractor.

In contrast if $x_0^2 \in (1, 1 + 2/\Delta t)$ then $|1 + \Delta t - \Delta t x_n^2| < 1$, then $\limsup_{n\to\infty} \|x_n\| \leqslant 1$, so that these trajectories do converge to the attractor $\mathcal{A} = [-1, 1]$ of (1.1.4). A general theory of local convergence for attractors has been developed, but we will not pursue it here; see [24] for details. We will see that when we add noise, all trajectories for the forward Euler method can escape to infinity no matter how small Δt is.

The previous example is a specific case of the forward Euler method applied to an equation satisfying Assumption 1.4.11. The behaviour seen, namely that some trajectories escape to infinity, happens for any f satisfying the same assumption.

1.5 Case study 1: The Langevin equation

Let $F : \mathbb{R}^d \to \mathbb{R}$, and consider the following ODE for $q, p \in \mathbb{R}^d$ the position and momenta of a particle of unit mass, namely

$$\dot{q} = p, \tag{1.5.1}$$

$$\dot{p} = -\gamma p - \nabla F(q). \tag{1.5.2}$$

The Langevin SDE, which we will consider later, is a generalisation of this ODE, which includes noise.

For (1.5.1), (1.5.2) setting $\gamma = 0$ gives a conservative Hamiltonian system, and the choice $\gamma > 0$ gives a damped Hamiltonian. To see this, note that, if we set

$$H(q, p) = \frac{1}{2}\|p\|^2 + F(q),$$

then

$$\dot{H} = \mathcal{L}H(q,p) = -\gamma\|p\|^2. \tag{1.5.3}$$

A generalisation of the Lyapunov function H will be useful in the following, and for this we require that $F(q) \geqslant 0$ for all $q \in \mathbb{R}^d$, and that there exists $\beta \in (0,1)$ and $\alpha \geqslant 0$ such that

$$\frac{1}{2}\langle \nabla F(q), q\rangle \geqslant \beta F(q) + \gamma^2 \frac{\beta(2-\beta)}{8(1-\beta)}\|q\|^2 - \alpha. \tag{1.5.4}$$

For example, if

$$F(q) = \frac{1}{4}(1 - \|q\|^2)^2, \tag{1.5.5}$$

then $F(q) \geqslant 0$ and, moreover, since

$$\nabla F(q) = -(1 - \|q\|^2)q,$$

we have

$$\langle \nabla F(q), q\rangle = (\|q\|^2 - 1)\|q\|^2, \tag{1.5.6}$$

and so the right and left hand sides of (1.5.4) are both quartics in $\|q\|$ and for any $\beta \in (0,1)$ and sufficiently large $\alpha > 0$ the inequality (1.5.4) holds.

Although H is non-increasing in time for $\gamma > 0$, its rate of change is zero whenever $p = 0$. In order to find a function whose rate of decrease is bounded from below outside a bounded set in phase space we introduce

$$V(q,p) = \frac{1}{4}\|p\|^2 + \frac{1}{4}\|p + \gamma q\|^2 + F(q). \tag{1.5.7}$$

Then

$$\begin{aligned}
\dot{V} &= \mathcal{L}V, \\
&= \frac{1}{2}\langle p, \dot{p}\rangle + \frac{1}{2}\langle \dot{p} + \gamma\dot{q}, p + \gamma q\rangle + \langle \nabla F(q), \dot{q}\rangle \\
&= -\frac{1}{2}\langle p, \gamma p + \nabla F(q)\rangle - \frac{1}{2}\langle \nabla F(q), p + \gamma q\rangle + \langle \nabla F(q), p\rangle \\
&= -\frac{\gamma}{2}\|p\|^2 - \frac{\gamma}{2}\langle \nabla F(q), q\rangle,
\end{aligned}$$

and using (1.5.4) gives

$$\dot{V} \leqslant -\gamma\left[\frac{1}{2}\|p\|^2 + \beta F(q) + \gamma^2\frac{\beta(2-\beta)}{8(1-\beta)}\|q\|^2\right] + \gamma\alpha.$$

But, for $\beta \in (0,1)$, straightforward calculation shows

$$\frac{1}{2}\|p\|^2 + \gamma^2\frac{\beta(2-\beta)}{8(1-\beta)}\|q\|^2 \geqslant \frac{\beta}{4}\|p\|^2 + \frac{\beta}{4}\|p + \gamma q\|^2.$$

Hence

$$\dot{V} \leqslant \gamma(\alpha - \beta V). \tag{1.5.8}$$

Thus applying the continuous Gronwall Lemma 1.2.10 we see that we approach any open neighbourhood of $\{x : V(x) \leqslant \alpha/\beta\}$ exponentially fast, and

$$\limsup_{t\to\infty} V(q(t), p(t)) \leqslant \frac{\alpha}{\beta}.$$

Thus by (1.5.7),

$$\limsup_{t\to\infty} \|p(t)\|^2 \leqslant \frac{4\alpha}{\beta}, \quad \text{and} \quad \limsup_{t\to\infty} \|p(t) + \gamma q(t)\|^2 \leqslant \frac{4\alpha}{\beta}.$$

Using $\|\gamma q\| \leqslant \|p + \gamma q\| + \|p\|$, we deduce further that

$$\limsup_{t\to\infty} \|q(t)\|^2 \leqslant \frac{16\alpha}{\gamma^2\beta}.$$

But we can say more. Because $H(q(t), p(t))$ is bounded below and monotonically decreasing, $\lim_{t\to\infty} H$ exists and

$$0 = \lim_{t\to\infty} \dot{H} = \lim_{t\to\infty} p(t) = \lim_{t\to\infty} \dot{q}(t),$$

using (1.5.3) and (1.5.1). It follows further that $\lim_{t\to\infty} \dot{p}(t) = 0$ and thus all solutions $(p(t), q(t))$ tend to the set of fixed points in the limit as $t \to \infty$ and, moreover, by (1.5.1)–(1.5.2) these fixed points satisfy

$$p = 0, \qquad \nabla F(q) = 0.$$

This problem is an example of a generalised gradient system, in the sense defined in [6]. The Lyapunov structure (1.5.8) is preserved by the backward Euler method—see [16]. In the next chapter we will study the effect of noise on this system.

1.6 Case study 2: particles in a velocity field

Consider the following model for particles moving in a two-dimensional velocity field:

$$\tau\ddot{x} = v(x) - \dot{x}. \tag{1.6.1}$$

Here $x \in \mathbb{T}^2$ (in this section $\mathbb{T}^2$ denotes the two-dimensional unit torus) denotes the particle position and $\dot{x} \in \mathbb{R}^2$ its velocity; we assume $\tau > 0$. The force on the particle is proportional to the difference between the fluid velocity at the particle site, $v(x)$, and the particle velocity, $\dot{x}$. The velocity field is given by $v = \nabla^\perp\psi$, where ψ is a stream-function, $\nabla^\perp$ denotes skew-gradient and hence $\nabla \cdot v = 0$. Thus x denotes the position of a particle moving according to Stokes' law in a two-dimensional incompressible velocity field v. Later on we will consider time-dependent random velocity fields $v(x, t)$.

Letting $y = \dot{x} \in \mathbb{R}^2$, we can write this as a system $\dot{u} = f(u)$ with $u = (x, y) \in \mathbb{X} = \mathbb{T}^2 \times \mathbb{R}^2$:

$$\begin{aligned} \dot{x} &= y, \\ \dot{y} &= \frac{1}{\tau}(v(x) - y). \end{aligned}$$

We assume that v is Lipschitz on $\mathbb{T}^2$. Then f is globally Lipschitz on $\mathbb{T}^2 \times \mathbb{R}^2$, and (1.6.1) defines a dynamical system.

We show that this dynamical system possesses a global attractor. The choice of $\mathbb{X} = \mathbb{T}^2 \times \mathbb{R}^2$ implies that $\|x(t)\| \leqslant 1$ for all $t \in \mathbb{R}$, and so it only remains to show that $\|y(t)\|$ is ultimately bounded independently of initial data to get an absorbing set. Now

$$\begin{aligned}\tau \frac{1}{2}\frac{d}{dt}\|y\|^2 &= \tau\langle y, \dot{y}\rangle \\ &= \langle v(x), y\rangle - \|y\|^2 \\ &\leqslant \frac{1}{2}[v(x)]^2 + \frac{1}{2}\|y\|^2 - \|y\|^2 \\ &\leqslant \frac{1}{2}c_v^2 - \frac{1}{2}\|y\|^2.\end{aligned}$$

Here $c_v = \max_{x\in\mathbb{T}^2}\|v(x)\| < \infty$, since a continuous function on a compact set achieves its supremum. Thus it follows that $\limsup_{t\to\infty}\|y(t)\| \leqslant c_v$ and so the system possesses an absorbing set and hence an attractor. On the attractor the particle velocity does not exceed the maximum fluid velocity in $\mathbb{T}^2$.

A prototypical example of a velocity field is given by the Taylor-Green explicit solution of the forced Navier-Stokes equations namely, for $x = (x_1, x_2)$,

$$\psi(x) = \sin(2\pi x_1)\sin(2\pi x_2).$$

Fig. 3 shows the direction field and streamlines induced by this velocity field; notice the four vortices. In the case $\tau = 0$, where particles follow fluid stream-lines, all solutions are periodic (except for a Lebesgue measure zero set of initial data lying on *heteroclinic orbits* connecting fixed points at $t = \pm\infty$) and there is no global attractor. For $\tau > 0$, however, there is a global attractor. Empirically we observe this to be made up of a finite number of periodic orbits, a finite number of fixed points, and orbits connecting them. This is illustrated in Fig. 4 which shows the distribution of 5000 particles, after sufficient time has elapsed for transient behaviour to disappear; what is seen is hence a visualisation of the global attractor in position space, excluding a number of heteroclinic orbits connecting pairs of periodic orbits and/or fixed points at $t = \pm\infty$.

The calculations are performed using the split-step Euler method

$$\begin{aligned}x_{n+1} &= x_n + \Delta t y_n, \\ y_{n+1} &= y_n + \frac{\Delta t}{\tau}(v(x_{n+1}) - y_{n+1}).\end{aligned}$$

This gives

$$\left(1 + \frac{\Delta t}{\tau}\right) y_{n+1} - \frac{\Delta t}{\tau} v(x_{n+1}) = y_n,$$

and so

$$\left(1 + \frac{\Delta t}{\tau}\right)^2 \|y_{n+1}\|^2 - 2\frac{\Delta t}{\tau}\left(1 + \frac{\Delta t}{\tau}\right)\langle y_{n+1}, v(x_{n+1})\rangle + \frac{\Delta t^2}{\tau^2}\|v(x_{n+1})\|^2 = \|y_n\|^2,$$

implying

$$\left(1 + \frac{\Delta t}{\tau}\right)^2 \|y_{n+1}\|^2 \leqslant \|y_n\|^2 + \frac{\Delta t}{\tau}\left(1 + \frac{\Delta t}{\tau}\right)[\|y_{n+1}\|^2 + \|v(x_{n+1})\|^2] - \frac{\Delta t^2}{\tau^2}\|v(x_{n+1})\|^2,$$

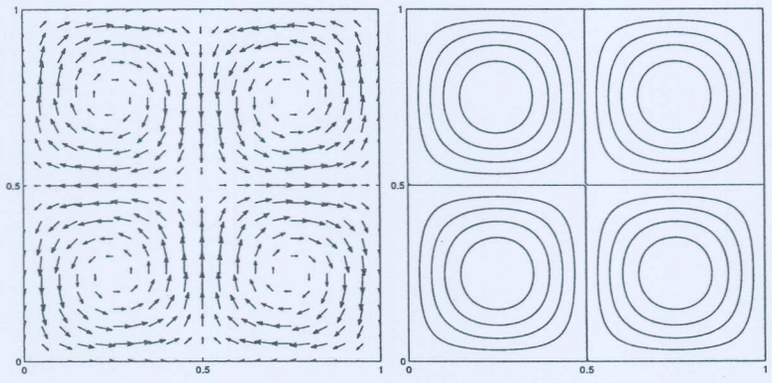


Figure 3: Taylor-Green velocity field: (i) direction field, (ii) streamlines.

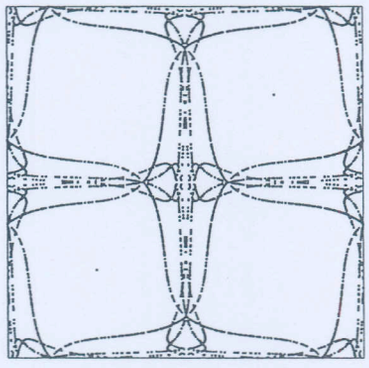


Figure 4: Global attractor for (1.6.1).

so that

$$\left(1+\frac{\Delta t}{\tau}\right)\|y_{n+1}\|^2 \leqslant \|y_n\|^2 + \frac{\Delta t}{\tau}\,c_v^2.$$

Thus, applying the discrete Gronwall Lemma 1.2.3, we obtain

$$\|y_n\|^2 \leqslant c_v^2 + \frac{1}{(1+\Delta t/\tau)^n}\,[\|y_0\|^2 - c_v^2],$$

and the existence of an absorbing set and global attractor follows. As for the exact solution, on the attractor the particle velocity does not exceed the maximum fluid velocity in $\mathbb{T}^2$.

2 Random dynamical systems

Here we take many of the ideas developed for dynamical systems and generalise them to situations where noise is present. This leads to the subject of random dynamical systems, recently given firm foundations in the book [1]. The subject can be rather technical upon first encounter. We aim to give an accessible introduction to some of it, in particular to those parts of relevance in numerical analysis. By being accessible we will miss out on many of the

subtleties of the subject and the reader is encouraged to study [1], and the references therein, for a thorough treatment.

2.1 Stochastic differential equations

Let $W(t)$ denote m-dimensional Brownian motion (see the discussion at the end of the section), $f : \mathbb{R}^d \to \mathbb{R}^d$ a smooth function and $\Sigma \in \mathbb{R}^{d\times m}$ a fixed matrix. Consider the following integral equation for $x \in C([0,T],\mathbb{R}^d)$:

$$x(t) = x(0) + \int_0^t f(x(s))\,ds + \Sigma W(t), \quad x(0) = X. \tag{2.1.1}$$

Underlying Brownian motion is *Wiener measure*, under which $W(t)$ is, with probability 1, in the space $C([0,T],\mathbb{R}^d)$ for any $T > 0$. However, also with probability 1, Brownian motion is nowhere differentiable. Thus it is not possible to differentiate (2.1.1) and find a non-autonomous differential equation in the usual sense. Nonetheless the equation is often written formally as the *stochastic differential equation* (SDE)

$$dx = f(x)dt + \Sigma dW, \quad x(0) = X. \tag{2.1.2}$$

If Σ depends upon t, explicitly and/or through $x(t)$, then it is necessary to introduce a notion of *stochastic integral* to interpret the term $\Sigma(t,x(t))dW(t)$ since, with probability 1, $W(t)$ is not of bounded variation. However, for the purpose of these notes we will consider the case of *additive noise* with Σ constant. The precise interpretation of (2.1.2) is then simply the integral equation (2.1.1).

Since $W(t)$ is, with probability 1, in $C([0,T],\mathbb{R}^d)$, it follows, under suitable conditions on f[1], that the solution $x(t)$ is itself a random function which is, with probability 1, in $C([0,T],\mathbb{R}^d)$. The objective of solving (2.1.2) is to find the properties of this random function, given the properties of Brownian motion. We denote Wiener measure by $\mathbb{P}$ and expectation under it by $\mathbb{E}$. The book [19] is a suitable introduction to the theory of SDEs; the books [8, 15] provide further useful development of the subject in the dynamical context of interest here.

Itô formula

In the following we assume that Σ is constant and either

(1) f is globally Lipschitz; or

(2) f is locally Lipschitz and satisfies a structural assumption inducing boundedness of solutions.

This is exactly what we did for ODEs. It is often of importance to understand how functions of $x(t)$ change with time. By using the *generator*:

$$\mathcal{L}V = \sum_{i=1}^{d} f_i \frac{\partial V}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d} [\Sigma\Sigma^T]_{ij} \frac{\partial^2 V}{\partial x_i \partial x_j}, \tag{2.1.3}$$

we may achieve this through the Itô formula [19]:

[1]The simplest being that f is globally Lipschitz; see [15] for generalisations.

2.1.1 Lemma *Let $x(t)$ solve (2.1.2) and let $V \in C^2(\mathbb{R}^d, \mathbb{R})$. Then the process $Y(t) = V(x(t))$ satisfies*

$$d\{V(x(t))\} = \mathcal{L}V(x(t))dt + \langle \nabla V(x(t)), \Sigma dW \rangle, \quad Y(0) = g(X).$$

To interpret this result we would need to introduce Itô stochastic integrals in order to make sense of the last term. However, for our purposes the following consequence is all we will need:

2.1.2 Lemma *Let $x(t)$ solve (2.1.2) and let $V \in C_0^2(\mathbb{R}^d, \mathbb{R})$ (C^2 functions with compact support). Then for $0 \leqslant s < t < \infty$ we have that $Y(t) = V(x(t))$ satisfies*

$$\mathbb{E}Y(t) = \mathbb{E}Y(s) + \mathbb{E}\int_s^t \mathcal{L}V(x(\tau))\, d\tau.$$

By means of *Dynkin's formula* [19] this result can be extended to $V \in C^2$ which are not compactly supported, provided sufficient boundedness of the solutions $x(t)$ is established. We do not provide details here, but freely apply Lemma 2.1.2 to non-compactly supported V when the necessary boundedness conditions on $x(t)$ hold.

A second (indirect) consequence of Lemma 2.1.2 is the *Fokker-Planck* equation for propagation of densities. For this we need the adjoint $\mathcal{L}^*$:

$$\mathcal{L}^*V = -\sum_{i=1}^d \frac{\partial}{\partial x_i}(f_i V) + \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j}\{[\Sigma\Sigma^T]_{ij} V\}.$$

2.1.3 Lemma *Assume that the law of $x(t)$, $\mathbb{P}\{x(t) \in A\}$, has a density $\rho(x,t) \in C^{2,1}(\mathbb{R}^d \times (0,T], \mathbb{R}^+)$. Then ρ satisfies the Fokker-Planck equation*

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^*\rho, \quad (x,t) \in \mathbb{R}^d \times (0,T],$$
$$\rho(x,0) = \delta(x - x_0), \quad x \in \mathbb{R}^d.$$

Numerical methods

We introduce three numerical methods for the SDE (2.1.2). The first, called the *Euler-Maruyama* method [11], is simply an SDE analogue of the *forward Euler* method (1.1.9) for deterministic systems. Given $t_n = n\Delta t$, and defining

$$\Delta W_n = W(t_{n+1}) - W(t_n),$$

the approximation x_n to $x(t_n)$ satisfies:

$$x_{n+1} = x_n + \Delta t f(x_n) + \Sigma \Delta W_n. \tag{2.1.4}$$

This may be derived, as in the deterministic case, by approximating the integral in (1.1.2). The second method considered we call the *stochastic backward Euler* method and is defined by

$$x_{n+1} = x_n + \Delta t f(x_{n+1}) + \Sigma \Delta W_n. \tag{2.1.5}$$

The final method, called the *split-step stochastic backward Euler* method, is defined by

$$\begin{aligned} x_* &= x_n + \Delta t f(x_*), \\ x_{n+1} &= x_* + \Sigma \Delta W_n. \end{aligned} \tag{2.1.6}$$

The idea of this method is similar to that underlying (1.1.11) with noise playing the role of f_2. For all three methods we choose $x_0 = X$ from (2.1.2).

Note that the Euler-Maruyama method is explicit, whereas the remaining two methods considered are implicit, requiring solution of a nonlinear equation for x_{n+1}, given x_n and ΔW_n. Here ΔW_n is an m-dimensional Gaussian random variable with mean 0 and covariance matrix $\Delta t I$; we denote this $\mathcal{N}(0, \Delta t I)$.

Brownian motion

Brownian motion is a continuous time analogue of the random walk. It is a process $\{W(t)\}_{t \geqslant 0}$, with $W(0) = 0$, and satisfying three basic properties:

(1) *time-homogeneity:* $W(t)$ and $W(t+s) - W(s)$ have the same distribution $\forall s, t \geqslant 0$;

(2) *independent increments:* $\{W(t_i) - W(s_i)\}_{i \geqslant 1}$ are independent random variables whenever the intervals $(t_i, s_i]$ are disjoint;

(3) *Gaussian increments:* $W(t) \sim \mathcal{N}(0, \sigma^2 t)$.

Standard Brownian motion denotes the choice $\sigma = 1$. Standard Brownian motion in $\mathbb{R}^m$ is a function $W : \mathbb{R}^+ \to \mathbb{R}^m$ with each component $W_i(t)$ being an independent standard Brownian motion.

It is possible to construct an underlying probability space in which such random functions live, and then to study the regularity of such functions under the resulting *Wiener measure*. See [9, 10] for details. Such analysis shows that, with probability 1, Brownian paths are α-Hölder continuous for any $\alpha \in [0, 1/2)$. This (lack of) regularity makes rates of convergence of approximations a more subtle issue than for ODEs.

Bibliographical remark The numerical analysis of SDEs is a subject in its infancy, though several books on the subject are now available, for example [2, 11]. Numerical analysis of SDEs in the context of random dynamical systems has not yet been developed in a systematic fashion and this fact will be manifest in our presentation. We will concentrate on two important dynamical concepts: (i) ergodicity and invariant measures and (ii) random attractors. Foundational work on the numerical analysis of SDEs according to their ability to reproduce ergodic properties is due to Talay [25, 26, 27]; see also [22, 16]. The study of random attractors, and the effect of discretisation, is being developed by Kloeden and co-workers (see [3, 12] for example), with recent extension to numerical approximation of SDEs by Robinson [20].

2.2 Random dynamical systems

We now generalise the concept of dynamical systems to allow for non-autonomous problems where the evolution in the space $\mathbb{X}$ depends upon time. Such time-dependence can be given

a variety of differing structures. Here we study *random dynamical systems* where the time-dependence is introduced through a *stationary* driving noise; by this we mean that the finite dimensional distributions of the random process are unaffected by translation in time. Our primary motivation is the study of the SDE (2.1.2) and its time-discrete counterparts, the forward and backward Euler methods (2.1.4), (2.1.5), (2.1.6). We start by giving the structure of the underlying noise model, and then show how to incorporate this into the definition of a random dynamical system. Our presentation is a curtailed version of [1, Chapter 1], which the reader should consult for greater depth and precision.

Noise model

A family $(\theta^t)_{t\in\mathbb{T}}$ of mappings of a measurable space $(\Omega, \mathcal{F})$ into itself is called a *measurable dynamical system* with time $\mathbb{T}$ if it satisfies the following conditions:

(1) $(\omega, t) \in (\Omega, \mathbb{T}) \mapsto \theta^t\omega \in \Omega$ is measurable;

(2) $\theta^0 = \mathrm{id}_\Omega$;

(3) $\theta^{s+t} = \theta^s \circ \theta^t \ \forall s, t \in \mathbb{T}$.

We will only consider the choices $\mathbb{T} = \mathbb{R}, \mathbb{Z}, \mathbb{R}^+, \mathbb{Z}^+$ for the time variable.

We now introduce a probability measure on the space $(\Omega, \mathcal{F})$ and consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that $\theta^t\mathbb{P} = \mathbb{P} \quad \forall t \in \mathbb{T}$: that is, $\mathbb{P}\{(\theta^t)^{-1}(A)\} = \mathbb{P}\{A\}$ for all $(t, A) \in (\mathbb{T}, \mathcal{F})$. The measure $\mathbb{P}$ is then said to be *invariant* with respect to θ. We have a *measure preserving* or *metric* dynamical system, which we denote by $(\Omega, \mathcal{F}, \mathbb{P}, \theta^t)$. This is the underlying noise model used in the theory of random dynamical systems. The noise is stationary, because of the invariance of $\mathbb{P}$ under θ.

Evolution driven by noise

In the following, $\mathcal{B}(\mathbb{T})$ denotes the Borel sets in $\mathbb{T}$.

2.2.1 Definition A *measurable random dynamical system* on a measurable space $(\mathbb{X}, V)$ over a metric dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, \theta^t)$ with time $\mathbb{T}$ is a mapping

$$\varphi : \mathbb{T} \times \mathbb{X} \times \Omega \to \mathbb{X}, \ (t, x, \omega) \mapsto \varphi(t, x, \omega),$$

where the mapping φ is $\mathcal{B}(\mathbb{T}) \otimes V \otimes \mathcal{F}$ measurable and satisfies the *co-cycle property* that $\varphi(0, x, \omega) = x$ for all $(x, \omega) \in (\mathbb{X}, \Omega)$ and, for all $t, s \in \mathbb{T}$, $x \in \mathbb{X}$, $\omega \in \Omega$,

$$\varphi(t + s, x, \omega) = \varphi(t, \varphi(s, x, \omega), \theta^s\omega).$$

It will sometimes be helpful to write

$$\varphi(t, x, \omega) = \varphi^t(\omega)x$$

as we will be interested in mapping sets in $\mathbb{X}$. In this notation the co-cycle property reads $\varphi^0(\omega) = \mathrm{id}_\mathbb{X}$ and

$$\varphi^{t+s}(\omega) = \varphi^t(\theta^s\omega) \circ \varphi^s(\omega).$$

φ is a generalisation of the evolution operator, now acting on noise as well as $\mathbb{X} \times \mathbb{T}$. We may think of such a process as a *skew-product* over the noise-process. A skew-product is a dynamical system such that

$$\begin{aligned}(Y, Z) &\to (Y, Z),\\ y &\mapsto f(y),\\ z &\mapsto g(y, z).\end{aligned}$$

The mapping f over Y is called the *base transformation*, and the mapping g over Z the *fibre transformation*.

In the case of random dynamical systems, the noise-process is the base, driving the evolution of real interest taking place in the fibre. We have $(Y, Z) = (\Omega, \mathbb{X})$ and define the measurable dynamical system Θ^t on $(\Omega \times \mathbb{X}, \mathcal{F} \otimes V)$ by

$$\Theta^t(x, \omega) = (\varphi^t(\omega)x, \theta^t\omega).$$

What distinguishes the random dynamical systems set-up from an arbitrary skew-product is the structure induced by $\mathbb{P}$. Without the stationarity assumption on the noise the framework would be too large to admit useful mathematical development; at the same time many problems arising in applications fit into the framework developed here. Together these two reasons suggest that developing the subject of random dynamical systems is a worthwhile enterprise.

Examples

We illustrate some of the concepts above by means of a number of simple examples. We will frequently invoke the *Markov property* [10]. Roughly, a random sequence or function $\{u_t\}_{t\in\mathbb{T}}$ is Markovian if the statistics of $u_t, t > s$ given u_s are independent of knowledge of u_τ for $\tau < s$. We will also frequently use the fact that a Gaussian measure is specified by its mean and covariance [10].

2.2.2 Example Consider a linearly damped map driven by white noise:

$$x_{n+1} = \lambda x_n + \eta_n,$$

where $|\lambda| < 1$ and the η_n form an independent identically distributed (i.i.d.) sequence of random variables with $\eta_0 \sim \mathcal{N}(0, \sigma^2)$.

In this case $\mathbb{T} = \mathbb{Z}^+$, $\mathbb{X} = \mathbb{R}$, $\Omega = \{\omega \in \mathbb{R}^{\mathbb{Z}^+} : \omega = (\eta_0, \eta_1, \dots)\}$ and θ^t is the shift on such sequences defined by $\theta^1\omega = (\eta_1, \eta_2, \dots)$. Then $\mathbb{P}$ is the measure induced on such i.i.d sequences by assuming $\eta_0 \sim \mathcal{N}(0, \sigma^2)$. The fact that the dynamical system for the noise is measure preserving follows since the η_j are i.i.d. The process $\{x_n\}$ is known as a discrete Ornstein-Uhlenbeck (OU) process.

2.2.3 Example We generalise the previous example to a linearly damped map driven by a coloured noise, namely the discrete-time OU process $\{\eta_n\}$:

$$\begin{aligned}x_{n+1} &= \lambda x_n + \eta_n,\\ \eta_{n+1} &= \gamma\eta_n + \xi_n,\end{aligned}$$

where $|\lambda|, |\gamma| < 1$ and ξ_n forms an i.i.d. sequence with $\xi_0 \sim \mathcal{N}(0,1)$. Again $\mathbb{T} = \mathbb{Z}^+$, $\mathbb{X} = \mathbb{R}$, $\Omega = \{\omega \in \mathbb{R}^{\mathbb{Z}^+} : \omega = (\eta_0, \eta_1, \dots)\}$ and θ^t is the shift map on sequences from the previous example. To ensure stationarity of the noise ω we choose η_0 independently of the i.i.d. sequence $\{\xi_j\}$ and set

$$\eta_0 \sim \mathcal{N}(0, \sigma^2),$$

where $\sigma^2 = (1-\gamma^2)^{-1}$. Then

$$\mathbb{E}\eta_1^2 = \gamma^2 \mathbb{E}\eta_0^2 + 1 = \mathbb{E}\eta_0^2.$$

By the Markov property the sequence ω is then stationary. To see the measure $\mathbb{P}$ explicitly we employ the formula

$$\eta_n = \gamma^n \eta_0 + \sum_{j=0}^{n-1} \gamma^j \xi_{n-1-j}.$$

The sequence $\{\eta_j\}$ is then seen to be Gaussian (since it is a linear combination of Gaussians) with mean zero and covariance

$$\mathbb{E}\eta_l \eta_m = \sigma^2 \gamma^{|m-l|}.$$

This completely specifies the invariant measure $\mathbb{P}$. Here we are using the fact that the mean and covariance define a Gaussian measure.

The preceding examples can be extended to $\mathbb{T} = \mathbb{Z}$ with bi-infinite noise sequences $\omega \in \mathbb{R}^{\mathbb{Z}}$. This extension will be necessary and relevant when studying random attractors.

2.2.4 Example Consider the SDE (2.1.2), together with the assumption that the drift f is a locally Lipschitz vector field satisfying

$$\exists \alpha > 0, \beta > 0 : \ \langle f(x), x \rangle \leqslant \alpha - \beta \|x\|^2. \tag{2.2.1}$$

Here $\mathbb{T} = \mathbb{R}^+$ and $\mathbb{X} = \mathbb{R}^d$. The set Ω comprises Brownian paths with invariant measure induced by the property of independent increments.[2]

Under our assumptions on f and Σ the solution of (2.1.2) exists for all $t > 0$ (essentially because the Lyapunov condition Assumption 2.3.6 holds for $V(x) = \|x\|^2$, see [8] or Theorem 36 in Chapter 2 of [15] for a precise statement of this result).

The preceding example can be extended to $\mathbb{T} = \mathbb{R}$ by considering Brownian motions with $t \in \mathbb{R}$. However, to do this requires choice of random initial data for $x(0)$ which is dependent on $\{W(t), t \leqslant 0\}$ [1]. The construction generalises the fact that, for deterministic dynamical systems, complete bounded orbits exist for all points on the global attractor [6, 24].

2.2.5 Example Consider the Euler-Maruyama method (2.1.4) for the approximation of (2.1.2). This can be formulated as a random dynamical system with $\mathbb{T} = \mathbb{R}^+$, $\mathbb{X} = \mathbb{R}^d$ and $\Omega = \{\omega \in \{\mathbb{R}^m\}^{\mathbb{Z}^+} : \omega = (\Delta W_0, \Delta W_1, \dots)\}$. Recall that $\{\Delta W_j\}$ forms an i.i.d. sequence so that, with θ being the shift defined in Example 2.2.2, we have the underlying measure preserving dynamical system representing the noise.

[2] More precisely, we work with equivalence classes of paths of the form $\{a + W(t), a \in \mathbb{R}^m\}$, where $W(t)$ is standard Brownian motion. See [1, Appendix A2] for details.

2.2.6 Example To pose the implicit methods (2.1.5), (2.1.6) as random dynamical systems it is necessary to prove existence and uniqueness for the implicit equations. This cannot be done in general, but can be achieved for certain natural structural assumptions, such as the one-sided Lipschitz condition Assumption 1.2.15. See [16] for details. Once this is done the formal set-up is similar to that of the previous example.

2.3 Invariant measures

The simplest way to introduce the concept of invariant measure is through the skew-product picture. Define $\pi_\Omega : \mathbb{X} \times \Omega \to \Omega$ by $\pi_\Omega(x, \omega) = \omega$. Notice that the structure of the skew-product means that

$$\pi_\Omega \circ \Theta^\bullet = \theta^\bullet \circ \pi_\Omega.$$

If μ is invariant for Θ, so that $\Theta^t \mu \equiv \mu$, then we see that

$$\pi_\Omega \circ \mu = \pi_\Omega \circ \Theta^t \circ \mu = \theta^t \circ (\pi_\Omega \circ \mu),$$

so that $\pi_\Omega \mu$ is $\theta(\cdot)$ invariant. Recalling that $\mathbb{P}$ is invariant for θ, these considerations motivate the following definition:

2.3.1 Definition Given a measurable random dynamical system φ, a probability measure μ on $(\mathbb{X} \times \Omega, V \times \mathcal{F})$ is an *invariant measure* for φ if

- $\Theta^t \mu = \mu \quad \forall t \in \mathbb{T}$;
- $\pi_\Omega \mu = \mathbb{P}$.

The measure μ will not be a product measure in general. However, the *disintegration*

$$\mu(A) = \int_A \mu_\omega(dx) \mathbb{P}(d\omega)$$

is often useful. If $\mu_\omega(dx)$ is independent of ω then the measure is product.

In all the examples we study here, the noise and the variable of interest which it drives will, together, form a Markov chain (discrete time) or Markov process (continuous time). We denote this Markov chain/process by (x_t, η_t) with $\eta_t \in \mathbb{Y}$ being the noise and $x_t \in \mathbb{X}$ being the noise driven process. Notice that ω is the *complete path* $\{\eta_t\}_{t \in \mathbb{T}} \in \Omega$. The preceding definition of invariant measure applies in the space $\mathbb{X} \times \Omega$. However, by virtue of the Markovian nature of our problems, it will be natural to seek invariant measures in $\mathbb{X} \times \mathbb{Y}$—to find a measure ν so that, if $(x_0, \eta_0) \sim \nu$ then $(x_t, \eta_t) \sim \nu$ for all $t \in \mathbb{T}$. Having found such a measure, additional statistics relating x_0 to $\{\eta_t\}_{t \in \mathbb{T}}$ are required to obtain the invariant measure μ in $\mathbb{X} \times \Omega$. For the first examples we will carry out this step of going from $\mathbb{X} \times \mathbb{Y}$ to $\mathbb{X} \times \Omega$. For the general theory of SDEs we will omit it. Indeed for such problems the measure ν is itself product measure in $\mathbb{X} \times \mathbb{Y}$, with the measure in $\mathbb{Y}$ being Gaussian measure on Brownian increments. So it is only necessary to look for an invariant measure in $\mathbb{X}$, and this is the viewpoint we take when looking at the general theory of ergodicity for SDEs.[3]

[3]Here we are implicitly assuming that $\mathbb{T} = \mathbb{R}^+$. In the case $\mathbb{T} = \mathbb{R}$ the invariant measure at $t = 0$ is correlated to the noise path for $t < 0$ [4].

2.3.2 Example Consider Example 2.2.2. We first find the invariant measure in $\mathbb{X}\times\mathbb{Y}$ where, here $\mathbb{X}=\mathbb{Y}=\mathbb{R}$. In fact the invariant measure in $\mathbb{X}\times\mathbb{Y}$ is product since the $\{\eta_n\}$ are i.i.d. To compute the invariant measure in the $\mathbb{X}$ coordinate, observe that $\{x_n\}$ is a Markov chain with Gaussian invariant measure $\mathcal{N}(0,\sigma^2/(1-\lambda^2))$. To see this we seek a Gaussian invariant measure in $\mathbb{X}$ with mean 0. Note that η_0 is independent of x_0 so that

$$\mathbb{E}|x_1|^2=\lambda^2\mathbb{E}|x_0|^2+\mathbb{E}|\eta_0|^2.$$

Equating $\mathbb{E}|x_1|^2$ and $\mathbb{E}|x_0|^2$ gives the variance of the invariant measure in $\mathbb{X}$. The invariant measure in $\mathbb{Y}$ is independent Gaussian $\mathcal{N}(0,\sigma^2)$.

We now lift this Gaussian invariant measure from $\mathbb{X}\times\mathbb{Y}$ to $\mathbb{X}\times\Omega$, seeking a Gaussian measure in

$$\{x_0,\{\eta_0,\eta_1,\eta_2,\dots\}\}\in\mathbb{X}\times\Omega.$$

We find that $\mathbb{E}x_0=\mathbb{E}\eta_j=0$, and

$$\mathbb{E}x_0^2=\frac{\sigma^2}{1-\lambda^2},\qquad \mathbb{E}\eta_j^2=\sigma^2$$
$$\mathbb{E}x_0\eta_i=0\quad\forall i,\qquad \mathbb{E}\eta_i\eta_j=0\quad\forall i\neq j.$$

We have verified the variance of x_0, the η_j are i.i.d. with the stated variance, and are therefore uncorrelated, whilst the condition on $\mathbb{E}x_0\eta_i$ follows provided x_0 is chosen independently of $\omega=(\eta_0,\eta_1,\dots)$. Because of this independence it follows that the disintegration yields product measure: $\mu_\omega(dx)$ is independent of ω.

2.3.3 Example Now we generalise this approach to finding the invariant measure to Example 2.2.3. Again we start in $\mathbb{X}\times\mathbb{Y}=\mathbb{R}\times\mathbb{R}$. The pairs $\{(x_n,\eta_n)\}$ form a Markov chain and its invariant measure is Gaussian with mean 0 and covariance matrix calculated as follows. We seek σ, μ and α:

$$\eta_0\sim\mathcal{N}(0,\sigma^2),\qquad x_0\sim\mathcal{N}(0,\mu^2),\qquad \mathbb{E}x_0\eta_0=\alpha.$$

Note that

$$\begin{aligned}\mathbb{E}\eta_1^2&=\gamma^2\mathbb{E}\eta_0^2+1\\ \mathbb{E}x_1\eta_1&=\lambda\gamma\mathbb{E}x_0\eta_0+\gamma\mathbb{E}\eta_0^2\\ \mathbb{E}x_1^2&=\lambda^2\mathbb{E}x_0^2+2\lambda\mathbb{E}x_0\eta_0+\mathbb{E}\eta_0^2.\end{aligned}$$

Invoking invariance (equating $\mathbb{E}\eta_1^2$ with $\mathbb{E}\eta_0^2$ and so forth) gives

$$\sigma^2=\frac{1}{(1-\gamma^2)},\qquad \alpha=\frac{\gamma\sigma^2}{(1-\gamma\lambda)},\qquad \mu^2=\frac{2\lambda\alpha+\sigma^2}{(1-\lambda^2)}.$$

Eliminating α from the expression for μ gives

$$\mu^2=\frac{(1+\gamma\lambda)\sigma^2}{(1-\gamma\lambda)(1-\lambda^2)}.$$

Thus we have an invariant measure on $\mathbb{X}\times\mathbb{Y}$. We now lift this to $\mathbb{X}\times\Omega$.

In the previous example x_0 was not correlated with $\omega = (\eta_0, \eta_1, \dots)$. In this example there is correlation and we must calculate this in order to completely specify the Gaussian invariant measure on $\mathbb{X} \times \Omega$ for the skew-product. If we let $a_j = \mathbb{E} x_0 \eta_j$, then straightforward calculation shows that

$$\begin{aligned} \mathbb{E} x_1 \eta_{j+1} &= \mathbb{E}[\lambda x_0 + \eta_0]\eta_{j+1} \\ &= \lambda \mathbb{E} x_0 \eta_{j+1} + \mathbb{E}\eta_0 \eta_{j+1}. \end{aligned}$$

Invoking invariance we see that

$$a_j = \lambda a_{j+1} + \sigma^2 \gamma^{j+1}.$$

In order to obtain a probability measure we need a bounded solution (as $j \to \infty$) and, since $|\lambda| < 1$, the only bounded solution of this recursion is

$$a_j = \mathbb{E} x_0 \eta_j = \frac{\sigma^2 \gamma^{j+1}}{(1 - \lambda\gamma)}.$$

In summary we have now found a Gaussian measure for

$$\{x_0, \{\eta_0, \eta_1, \eta_2, \dots\}\} \in \mathbb{X} \times \Omega$$

which is specified by $\mathbb{E} x_0 = \mathbb{E}\eta_j = 0$ and

$$\mathbb{E} x_0^2 = \mu^2, \qquad \mathbb{E}\eta_j^2 = \sigma^2$$
$$\mathbb{E} x_0 \eta_i = a_j \quad \forall i, \qquad \mathbb{E}\eta_i \eta_j = \sigma^2 \gamma^{|i-j|} \quad \forall i \neq j.$$

For this problem the disintegration is not that of product measure, due to the correlation of x_0 with $\omega = (\eta_0, \eta_1, \dots)$, manifest in non-zero a_j.

In both of these examples we have calculated the invariant measure by finding the invariant measure of a Markov chain on $\mathbb{X} \times \mathbb{Y}$, and then finding additional correlations between the solution in the fibre and the noise in the base space; this then gives the invariant measure on $\mathbb{X} \times \Omega$. This approach can also be used for studying SDEs, and their time discretisations, using a Markov process in the case of the SDE. Because the calculation of correlations between the fibre and base solutions is straightforward, but tedious, we will omit it in the developments which follow Assumption 2.3.6. The essence of the calculation is to find the invariant measure for the underlying Markov process itself in $\mathbb{X} \times \mathbb{Y}$, and we concentrate on this in our examples of SDEs.

For SDEs we note that Lemma 2.1.3 holds the key to understanding the existence of invariant measures. We are interested in finding conditions under which there is a unique solution $\bar{\rho}(x)$ of

$$\mathcal{L}^* \bar{\rho} = 0, \qquad \int_{\mathbb{R}^d} \bar{\rho}(x)\, dx = 1 \tag{2.3.1}$$

and that ρ converges to $\bar{\rho}$ in the weak-$*$ topology (identifying ρ with the measure it induces). This means that for a suitable class of functions g,

$$\int_{\mathbb{R}^d} \rho(x,t) g(x)\, dx \to \int_{\mathbb{R}^d} \bar{\rho}(x) g(x)\, dx$$

as $t \to \infty$ or, in more abstract notation,

$$\mathbb{E}g(x(t)) \to \pi(g),$$

where π is the measure with density $\bar{\rho}$.

2.3.4 Example We generalise Example 2.3.2 to continuous time by studying the Ornstein-Uhlenbeck (OU) process

$$dx = -xdt + \sigma dW, \quad x(0) = x_0. \tag{2.3.2}$$

Thus $\mathbb{X} = \mathbb{R}$, $\mathbb{T} = \mathbb{R}^+$. A straightforward calculation, using the representation of the solution as an Itô stochastic integral, shows that

$$x(t) \sim \mathcal{N}(e^{-t}y_0, \sigma^2[1 - e^{-2t}]/2),$$

indicating convergence to the Gaussian invariant measure $\mathcal{N}(0, \sigma^2/2)$ as $t \to \infty$ in $\mathbb{X}$. The invariant measure in $\mathbb{Y}$ (where the driving Brownian motion lies) is independent of that in $\mathbb{X}$ and we do not discuss it explicitly.

It is readily verified that the density associated with the Gaussian measure for $x(t)$ is a steady state solution of the Fokker-Planck equation of Lemma 2.1.3 which reduces to

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(-x\rho) = \frac{\sigma^2}{2}\frac{\partial^2 \rho}{\partial x^2}, \quad \rho(x, 0) = \delta(x).$$

2.3.5 Example Now we generalise the Example 2.3.3 to continuous time. Consider the equations

$$\begin{aligned} dx &= [-ax + \eta]dt, \quad x(0) = x_0 \\ d\eta &= -\eta dt + \sigma dW, \quad \eta(0) = \eta_0. \end{aligned}$$

We assume $a > 0$. Here $\mathbb{X} = \mathbb{Y} = \mathbb{R}$, $\mathbb{T} = \mathbb{R}^+$. We are viewing η as the driving noise, not W.

The Fokker-Planck equation takes the form

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}\{[-ax + \eta]\rho\} + \frac{\partial}{\partial \eta}\{-\eta\rho\} = \frac{\sigma^2}{2}\frac{\partial^2 \rho}{\partial \eta^2}.$$

The previous example shows that η has Gaussian distribution and converges to a Gaussian invariant measure. Since

$$x(t) = e^{-at}x(0) + \int_0^t e^{-a(t-s)}\eta(s)\, ds$$

and η is Gaussian we deduce that x too is Gaussian. Note, however, that it is correlated to η, unlike the previous example.

These considerations suggest that we seek a steady solution of the Fokker-Planck equation in the form

$$\bar{\rho}(x, \eta) \propto \exp\{-\alpha x^2 + \beta x\eta - \gamma\eta^2\}.$$

(The constant of proportionality should be chosen so that $\bar{\rho}$ integrates to 1 on $\mathbb{R}^2$.) Substitution shows that

$$\alpha = \frac{a(a+1)^2}{\sigma^2}, \qquad \beta = \frac{2a(a+1)}{\sigma^2}, \qquad \gamma = \frac{(a+1)}{\sigma^2}$$

and the Gaussian invariant measure for $(x, \eta) \in \mathbb{X} \times \mathbb{Y}$ is found; note that x and η are correlated as $\beta \neq 0$.

To find the invariant measure for the skew-product $\{x_0, \{\eta(t)\}_{t\in\mathbb{R}^+}\} \in \mathbb{X}\times\Omega = \mathbb{R}\times\mathbb{R}^{\mathbb{R}^+}$, it is necessary to calculate how x_0 correlates with the entire path $\{\eta(t)\}_{t\in\mathbb{R}^+}$. A straightforward, but tedious, calculation shows that

$$\mathbb{E}x(t)\eta(t+\xi) = \frac{\sigma^2}{2(1+a)}\, e^{-\xi} \quad \forall t, \xi \in \mathbb{R}^+.$$

We now describe an abstract theory which is useful for finding ergodic invariant measures for SDEs where explicit solutions of the Fokker-Planck equation are not known. We look only for invariance in $\mathbb{X}$, viewing the Brownian motion evolving in $\mathbb{Y}$, with complete path in Ω, as the noise.

The following two assumptions encode two basic properties needed for ergodicity: *tightness* to ensure that probability does not leak out to infinity, and *reachability* showing that neighbourhoods of every point in phase space can be reached by appropriate choice of noise. There are two versions of the Lyapunov condition, one for continuous time $\mathbb{T} = \mathbb{R}$ or $\mathbb{R}^+$, and one for discrete time $\mathbb{T} = \mathbb{Z}$ or $\mathbb{Z}^+$.

2.3.6 Assumption (Lyapunov Condition—Continuous Time) There is a function $V : \mathbb{X} \to [1, \infty)$, with $\lim_{\|x\|\to\infty} V(x) = \infty$, and real numbers $a \in (0, \infty)$, $d \in (0, \infty)$ such that

$$\mathcal{L}\{V(x)\} \leqslant -a\{V(x)\} + d, \tag{2.3.3}$$

where $\mathcal{L}$ is the generator for (2.1.2) given by (2.1.3).

Let $\mathcal{F}_n$ denote the sub σ-algebra of all events up to time n (see [15]).

2.3.7 Assumption (Lyapunov Condition—Discrete Time) There is a function $V : \mathbb{X} \to [1, \infty)$, with $\lim_{\|x\|\to\infty} V(x) = \infty$, and real numbers $\alpha \in (0, 1)$, and $\beta \in [0, \infty)$ such that

$$\mathbb{E}[V(x_{n+1})|\mathcal{F}_n] \leqslant \alpha V(x_n) + \beta.$$

2.3.8 Assumption The Markov chain or process $\{x_t\}$, $t \in \mathbb{T}$, with transition kernel $P_t(x, A)$ satisfies, for some fixed compact set $C \in \mathcal{B}(\mathbb{X})$, the following:

(1) (*reachability*) for some $y^* \in \text{int}(C)$ and for any $\delta > 0$, there is a $t_1 = t_1(\delta) \in \mathbb{T}$ such that

$$P_{t_1}\Big(x, B(y^*, \delta)\Big) > 0 \quad \forall x \in C;$$

(2) (*smoothness*) for $t \in \mathbb{T}$ the transition kernel possesses a density $p_t(x, y)$, precisely

$$P_t(x, A) = \int_A p_t(x, y)\, dy \quad \forall x \in C,\ A \in \mathcal{B}(\mathbb{X}) \cap \mathcal{B}(C),$$

and $p_t(x, y)$ is jointly continuous in $(x, y) \in C \times C$.

In what follows, we will use the shorthand notation $\|f\| \leqslant V$ to mean $\|f(x)\| \leqslant V(x)$ for all $x \in \mathbb{X}$, and define

$$\mathcal{G} = \{\text{measurable } g : \mathbb{X} \to \mathbb{R} \text{ with } \|g\| \leqslant V\}.$$

The following ergodic theorem, which follows from a straightforward development of ideas in [8, 18], will give invariant measures for a variety of SDEs and their discretisations.

2.3.9 Theorem *Let $x(t)$ denote the solution of the SDE* (2.1.2) *(resp. a discrete time Markov chain) with transition kernel $P_t(x, A)$. Assume that there is a $T > 0$ for which the following holds: the Markov process (resp. chain) satisfies Assumptions* 2.3.6 *(resp.* 2.3.7*) and* 2.3.8 *with C given by*

$$C = \left\{ x : V(x) \leqslant \frac{2\beta}{\gamma - \alpha} \right\}$$

for some $\gamma \in (\alpha^{1/2}, 1)$. Then there exists a unique invariant measure π. Furthermore, there is $\mu(\gamma) \in (0,1)$ and $\kappa(\gamma) \in (0, \infty)$ such that for all measurable $g \in \mathcal{G}$

$$|\mathbb{E}^{x_0} g(x(t)) - \pi(g)| \leqslant \kappa e^{-\mu t} V(x_0).$$

Proof See [16]. □

2.3.10 Example Consider Example 2.2.4 which generates a random dynamical system on $\mathbb{X} = \mathbb{R}^d$. If $V(x) = \frac{1}{2}\|x\|^2 + 1$ then (2.2.1) shows that

$$\mathcal{L}V(x) = -2\beta V(x) + \alpha + \beta + \frac{1}{2}\sigma^2,$$

where $\sigma^2 = \|\Sigma\|_F^2$ and $\|\cdot\|_F$ denotes the Frobenius norm. Thus Assumption 2.3.6 holds. Assumption 2.3.8 may also be established easily if $m \geqslant d$, and we assume this here. By Theorem 2.3.9 the SDE then has a unique exponentially attracting invariant measure on $\mathbb{R}^d$ [16]; this can be lifted to an invariant measure for the skew-product on $\mathbb{R}^d \times \Omega$ but we omit the details.

The harder case $m < d$ is also studied in [16]. It is harder because the Fokker-Planck equation of Lemma 2.1.3 is no longer uniformly parabolic—there is diffusion only in m directions for a PDE in d space dimensions. This makes both smoothness and reachability impossible to establish in general: it is necessary to study the interaction of the deterministic flow (manifest in the hyperbolic part of the Fokker-Planck equation) with the noise (manifest in the directions in which parabolic behaviour is found) to find when they hold. See [16] for examples where this can be done, and for references to the literature.

2.3.11 Example Consider the SDE

$$dx = -x^3 dt + dW, \quad x(0) = X.$$

This equation satisfies (2.2.1) and is hence geometrically ergodic. In particular, for all polynomials V,

$$\mathbb{E}^{x_0} V(x(t)) \to \pi(g),$$

exponentially fast in time. In fact π is explicitly known and has density proportional to $\exp(-x^4/2)$, as can be verified from the Fokker-Planck equation. However, the Euler-Maruyama scheme (2.1.4) is not ergodic for any $\Delta t > 0$, however small. In particular, for any $\Delta t > 0$ there is x_0 such that $\mathbb{E}|x_n|^2 \to \infty$ as $n \to \infty$. Furthermore, $\mathbb{P}\{\limsup_{n\to\infty} |x_n| = \infty\} > 0$ for any $\Delta t > 0$ and any $x_0 \in \mathbb{R}^d$. (See [16] and references to related results contained therein).

This is the random analogue of our observations in Example 1.4.18. The situation here is much more severe, however, since for *any* $\Delta t > 0$, $x_0 \in \mathbb{R}^d$ there is positive probability of blow-up. For the deterministic case blow-up is avoided for fixed x_0 by reduction of Δt.

2.3.12 Example To overcome the difficulties of the previous example, consider the split-step backward Euler method (2.1.6) applied to (2.1.2) under (2.2.1) which implies Assumption 1.4.11 with $p = 1$. Since

$$x_* - \Delta t f(x_*) = x_n$$

it follows that

$$\begin{aligned} \|x_*\|^2 &= \|x_n\|^2 + 2\Delta t\langle x_*, f(x_*)\rangle - \Delta t^2\|f(x_*)\|^2 \\ &\leqslant \|x_n\|^2 + 2\Delta t(\alpha - \beta\|x_*\|^2). \end{aligned}$$

Rearranging we have that

$$\|x_*\|^2 \leqslant \frac{2\Delta t\alpha}{1+2\Delta t\beta} + \frac{1}{1+2\Delta t\beta}\|x_n\|^2,$$

and hence that

$$\|x_{n+1}\|^2 \leqslant \frac{2\Delta t\alpha}{1+2\Delta t\beta} + \frac{1}{1+2\Delta t\beta}\|x_n\|^2 + 2\langle x_*, \Sigma\Delta W_n\rangle + \|\Sigma\Delta W_n\|^2.$$

Now the ΔW_n are i.i.d. random variables with

$$\mathbb{E}\|\Sigma\Delta W_n\|^2 = \sigma^2\Delta t.$$

Thus

$$\mathbb{E}(\|x_{n+1}\|^2|\mathcal{F}_n) \leqslant \frac{1}{1+2\Delta t\beta}\|x_n\|^2 + \frac{2\Delta t\alpha}{1+2\Delta t\beta} + \sigma^2\Delta t.$$

This follows since x_* is independent of ΔW_n so that

$$\mathbb{E}(\langle x_*, \Sigma\Delta W_n\rangle|\mathcal{F}_n) = 0.$$

We have established that Assumption 2.3.7 (the Lyapunov condition) holds for this numerical method. If $m = d$ then, as for the SDE itself, reachability and smoothness are straightforward. Hence by Theorem 2.3.9, the numerical method is ergodic. Problems with $m < d$ can be studied similarly. For details see [16].

2.4 Random attractors

In the previous sections we generalised deterministic invariant sets to consider invariant measures for random dynamical systems. Attractors can also be usefully generalised to random dynamical systems. To do this it is important to observe that $\varphi(t, \theta^{-t}\omega)x_0$ denotes the solution at time 0 starting from x_0 at time $-t$. A random attractor is defined as a set $\mathcal{A}$ such that $\varphi(t, \theta^{-t}\omega)B$ approaches $\mathcal{A}$ as $t \to \infty$ for all bounded B. It has the interpretation as the set in which all solutions will lie at time $t = 0$ when the initial data at time $t = -\infty$ is in a bounded set. This kind of convergence, involving initial data at time $-\infty$, is known as *pull-back* convergence. Because of this, it is necessary in this section to consider $\mathbb{T} = \mathbb{R}$ or $\mathbb{Z}$ (and not $\mathbb{R}^+$ or $\mathbb{Z}^+$). In the deterministic case, where $\varphi = \varphi(t)$ only, the definition of random attractor coincides with the deterministic definition (Definition 1.4.3) of an attractor.

A *random closed (compact) set* is a mapping $A : \Omega \to V$, where $A(\omega)$ is closed (compact) for all $\omega \in \Omega$ and the mapping $\omega \mapsto \text{dist}(x, A(\omega))$ is measurable for each $x \in \mathbb{X}$. The mapping $U : \Omega \to V$ is a *random open* set if $U^c : \Omega \to V$ is a random closed set.

2.4.1 Definition A random set K is a *random absorbing set* if, for all bounded $B \subset X$ and with probability one, there exists $t^* = t^*(B, \omega) > 0$ such that

$$\varphi(t, \theta^{-t})B \subset K(\omega) \quad \forall t \geqslant t^*.$$

2.4.2 Definition The set $\mathcal{A}(\omega) \subset X$ is a *random global attractor* if, with probability one,

- $\mathcal{A}(\omega)$ is a random compact set;
- $\varphi(t, \omega)\mathcal{A}(\omega) = \mathcal{A}(\theta^t\omega)$ $\forall t \geqslant 0$ (invariance);
- for bounded non-random B

$$\lim_{t\to\infty} \text{dist}(\varphi(t, \theta^{-t}\omega)B, \mathcal{A}(\omega)) = 0,$$

  where dist is defined in Definition 1.4.1.

The following theorem from [1] is valid in both discrete and continuous time, with $t, T \in \mathbb{T}$.

2.4.3 Theorem [4] *If a random dynamical system has a random absorbing set, then*

$$\mathcal{A}(\omega) := \overline{\bigcup_{B\subset X} \Lambda(B, \omega)}$$

is a random global attractor, where the union is over all bounded $B \subset X$, and

$$\Lambda(B, \omega) = \bigcap_{T\geqslant 0} \overline{\bigcup_{t\geqslant T} \varphi(t, \theta^{-t}\omega)B}$$

is the ω-limit set *of B.*

The notion of attractor based on data at time $-\infty$ is not always what is needed in application. Hence the following is of interest:

2.4.4 Theorem [4] *If $\mathcal{A}$ is a random global attractor, then*

$$\operatorname{dist}(\varphi(t,\omega)B, \mathcal{A}(\theta^t\omega)) \to 0$$

in probability.

2.4.5 Example For both the Examples 2.2.2, 2.2.3 we have

$$x_{l+m} = \lambda^m x_l + \sum_{j=0}^{m-1} \lambda^j \eta_{l+m-1-j}.$$

For both these examples the random attractor is a (random) point. To see the pull-back convergence, set $l + m = 0$ and $x_l = y$ to obtain

$$x_0 = \lambda^m y + \sum_{j=0}^{m-1} \lambda^j \eta_{-1-j}.$$

Letting $m \to \infty$ gives a natural guess for the random attractor, namely

$$\mathcal{A}(\omega) = \sum_{j=0}^{\infty} \lambda^j \eta_{-1-j},$$

and this infinite sum can be shown to converge, with probability one, for both examples. Furthermore, convergence of x_0 to $\mathcal{A}(\omega)$ with probability one can also be shown. Thus the random attractor is a point for each $\omega \in \Omega$.

To illustrate the forward convergence we set $l = 0$, $x_0 = y$ to obtain

$$x_m = \lambda^m y + \sum_{j=0}^{m-1} \lambda^j \eta_{m-1-j}.$$

Now

$$\mathcal{A}(\theta^m\omega) = \sum_{j=0}^{\infty} \lambda^j \eta_{m-1-j}.$$

Hence

$$\begin{aligned} |x_m - \mathcal{A}(\theta^m\omega)| &= |\lambda^m y + \sum_{j=m}^{\infty} \lambda^j \eta_{m-1-j}| \\ &= |\lambda^m| \cdot |y + \sum_{j=m}^{\infty} \lambda^{j-m} \eta_{m-1-j}| \\ &= |\lambda|^m |y + \mathcal{A}(\omega)| \\ &\to 0 \end{aligned}$$

as $m \to \infty$ with probability 1. Thus, in this particular case, the forward convergence to the random attractor is almost sure; this is stronger than the convergence in probability guaranteed by Theorem 2.4.4 and is a consequence of the simple (point) structure of the random attractor here.

Now we study the SDE (2.1.2), together with the two implicit backward Euler approximations of it. We employ Dissipativity Assumption 1.4.11. It is straightforward to see that, for continuous f, this implies that there exist $\alpha_0, \beta_0, \gamma_0, \eta_0 > 0$ and $p \geqslant 1$ such that

$$\begin{aligned} \langle x, f(x)\rangle &\leqslant \alpha_0 - \beta_0\|x\|^{2p} \qquad \forall x \in \mathbb{R}^d, \\ \|f(x)\| &\leqslant \eta_0 + \gamma_0\|x\|^{2p-1} \quad \forall x \in \mathbb{R}^d. \end{aligned}$$

Note that the first of these expressions implies that there are $\alpha_1, \beta_1 > 0$ such that

$$\langle x, f(x)\rangle \leqslant \alpha_1 - \beta_1\|x\|^2 - \frac{1}{2}\beta_0\|x\|^{2p} \quad \forall x \in \mathbb{R}^d. \tag{2.4.1}$$

A central lemma for both the continuous and discrete time cases is:

2.4.6 Lemma *Under Assumption* 1.4.11 *there exists* $C > 0$ *such that*

$$I := \langle f(v+z) + \beta_1(v+z), v\rangle \leqslant C[1 + \|z\|^2 + \|z\|^{2p}].$$

Proof From Assumptions 1.4.11 (and the implied equation (2.4.1))

$$\begin{aligned} I &= \langle f(v+z) + \beta_1(v+z), v+z\rangle - \langle f(v+z) + \beta_1(v+z), z\rangle \\ &\leqslant \alpha_1 - \frac{1}{2}\beta_0\|v+z\|^{2p} + \|f(v+z)\|\,\|z\| + \beta_1\|v+z\|\,\|z\| \\ &\leqslant \alpha_1 - \frac{1}{2}\beta_0\|v+z\|^{2p} + \eta_0\|z\| + \gamma_0\|v+z\|^{2p-1}\|z\| + \frac{\beta_1\delta^2}{2}\|v+z\|^2 + \frac{\beta_1}{2\delta^2}\|z\|^2. \end{aligned}$$

By the Young inequality, for all $\varepsilon > 0$,

$$\|v+z\|^{2p-1}\|z\| \leqslant \varepsilon\|v+z\|^{2p} + \frac{1}{2p\varepsilon^{2p-1}}\|z\|^{2p}$$

and so, by choosing ε and then δ sufficiently small, the result follows. □

2.4.7 Example We now outline how to prove the existence of a random global attractor for the SDE; the method will be mimicked for the backward Euler methods. First let $z(t), t \in \mathbb{R}$ be the stationary solution of the continuous time OU process

$$dz + \beta_1 z dt = \Sigma dW \tag{2.4.2}$$

whose solution we studied in Example 2.3.4 in the case $d = 1$.

An important observation in what follows is that z has growth which is at most logarithmic in $t \to \pm\infty$ [15]: with probability one there are constants $A, B > 0$ such that

$$\|z(t)\|^2 \leqslant A + B\log[1 + |t|]. \tag{2.4.3}$$

This generalises the law of the iterated logarithm from Brownian motion to the OU process. It is proved using the exponential Martingale inequality (see [15, Chapter 1, Theorem 7.4]). If we define $v := x - z$ then

$$\frac{dv}{dt} + \beta_1 v = f(v+z) + \beta_1(v+z). \tag{2.4.4}$$

We will show that v, and hence x, has a random absorbing set, leading to the existence of the random global attractor. The motivation for introducing v is that it satisfies a standard ODE with random right hand side, rather than an SDE. It is hence easier to directly obtain uniform in t bounds for v than for x.

2.4.8 Theorem *Let the Dissipativity Assumption* 1.4.11 *hold. Then x solving* (2.1.2) *generates a random dynamical system on $\mathbb{R}^d$ with $\mathbb{T} = \mathbb{R}$ and has a random global attractor.*

Proof The existence and uniqueness of solutions follows by application a Theorem 3.6 in Chapter 2 of [15], after noting that (2.4.1) implies that

$$\langle x, f(x)\rangle \leqslant \alpha_1 - \beta_1\|x\|^2 \quad \forall x \in \mathbb{R}^d. \tag{2.4.5}$$

By Lemma 2.4.6 we have, from (2.4.4),

$$\frac{1}{2}\frac{d}{dt}\|v\|^2 + \beta_1\|v\|^2 \leqslant C[1 + \|z\|^2 + \|z\|^{2p}].$$

Thus

$$\begin{aligned}\|v(0)\|^2 - e^{-2\beta_1 t}\|v(-t)\|^2 &\leqslant \int_{-t}^{0} Ce^{2\beta_1 s}[1 + \|z(s)\|^2 + \|z(s)\|^{2p}]\,ds \\ &\leqslant \int_{-\infty}^{0} Ce^{2\beta_1 s}[1 + \|z(s)\|^2 + \|z(s)\|^{2p}]\,ds \\ &= R(\omega).\end{aligned}$$

Now

$$\begin{aligned}\|x(0)\|^2 &\leqslant 2\|v(0)\|^2 + 2\|z(0)\|^2 \\ &\leqslant 4e^{-2\beta_1 t}[\|x(-t)\|^2 + \|z(-t)\|^2] + 2R(\omega) + 2\|z(0)\|^2.\end{aligned}$$

Note that, by (2.4.3), almost surely

$$R(\omega) < \infty.$$

Thus if $\|x(-t)\|^2 \leqslant \rho$ then, almost surely, there exists $t^*(\rho,\omega) > 0$ such that $\|x(0)\|^2 \leqslant r(\omega) < \infty$, where

$$r(\omega) := 2R(\omega) + 2\|z(0)\|^2 + \varepsilon$$

for any $\varepsilon > 0$. By Theorem 2.4.3 the proof is complete. □

2.4.9 Example We now move on to consider the approximation schemes (2.1.5) and (2.1.6). In addition to the Dissipativity Assumption 1.4.11, we also assume the one-sided Lipschitz condition (1.2.6). This enables us to prove existence and uniqueness of the random dynamical system generated by these two methods, for all Δt sufficiently small.

2.4.10 Theorem *Let Assumptions* 1.4.11 *and* 1.2.15 *hold. Then both the backward Euler method* (2.1.5) *and the split-step backward Euler method* (2.1.6) *define random dynamical systems on $\mathbb{R}^m$ with $\mathbb{T} = \mathbb{Z}$ and have a random global attractor.*

Proof The existence of a random dynamical system follows from Lemma 1.2.19 for both methods. For the backward Euler method we have

$$(1 + \Delta t\beta_1)x_{n+1} = x_n + \Delta t[f(x_{n+1}) + \beta_1 x_{n+1}] + \Sigma\Delta W_n.$$

Define Z_n to be the discrete OU process

$$(1 + \Delta t\beta_1)Z_{n+1} = Z_n + \Sigma\Delta W_n$$

and $V_n := x_n - Z_n$. Then

$$(1 + \Delta t\beta_1)V_{n+1} = V_n + \Delta t[f(x_{n+1}) + \beta_1 x_{n+1}],$$

and hence, using Lemma 2.4.6,

$$\begin{aligned}(1 + \Delta t\beta_1)\|V_{n+1}\|^2 &\leqslant \langle V_n, V_{n+1}\rangle + C[1 + \|Z_{n+1}\|^2 + \|Z_{n+1}\|^{2p}] \\ &\leqslant \frac{1}{2}\|V_n\|^2 + \frac{1}{2}\|V_{n+1}\|^2 + C[1 + \|Z_{n+1}\|^2 + \|Z_{n+1}\|^{2p}].\end{aligned}$$

Re-arranging gives

$$(1 + 2\Delta t\beta_1)\|V_{n+1}\|^2 \leqslant \|V_n\|^2 + 2C[1 + \|Z_{n+1}\|^2 + \|Z_{n+1}\|^{2p}].$$

Using the discrete Gronwall Lemma 1.2.11 gives the desired absorbing set, and hence attractor, after using a discrete time analogue of (2.4.3); this can be drived by use of the exponential Martingale inequality (see [15, Theorem 7.4, Chapter 1]).

For the split-step backward Euler method (2.1.6),

$$\begin{aligned}(1 + \Delta t\beta_1)x_* &= x_n + \Delta t[f(x_*) + \beta_1 x_*] \\ x_{n+1} &= x_* + \Sigma\Delta W_n.\end{aligned}$$

Define Z_n by

$$\begin{aligned}(1 + \Delta t\beta_1)Z_* &= Z_n \\ Z_{n+1} &= Z_* + \Sigma\Delta W_n.\end{aligned}$$

If $V_n := x_n - Z_n$ and $V_* := x_* - Z_*$ then

$$(1 + \Delta t\beta_1)V_* = V_n + \Delta t[f(V_* + Z_*) + \beta_1(V_* + Z_*)].$$

But $V_* = V_{n+1}$ and so the analysis now proceeds as for the backward Euler method, replacing Z_{n+1} by $Z_* = (1 + \Delta t\beta_1)^{-1}Z_n$. □

2.5 Case study 1: the Langevin equation

The Langevin equation plays a central role in statistical physics. It describes how a mechanical system, with Hamiltonian $\frac{1}{2}\|p\|^2 + F(q)$, behaves when it is placed in contact with a heat bath: a larger Hamiltonian system. The overall coupled system is assumed Hamiltonian and then a variety of arguments, most using the large relative size of the heat bath, lead to an equation for p, q which includes the effect of energy exchange with the heat bath. This energy exchange is in the form of damping, from (p, q) to the bath, and noise which adds energy from the bath to (p, q), in mean square. The paper [5] gives a derivation of the equation from a simple mechanical system. Other derivations are less precise [29], invoking empirical assumptions about the effect of coupling, but have led to a model which is useful in many contexts.

Let $W(t)$ be standard d-dimensional Brownian motion, $F : \mathbb{R}^d \to \mathbb{R}$, $\sigma \in \mathbb{R}^{d\times d}$ and $\rho_i \in \mathbb{R}^d$ be the ith column of σ; we assume that the ρ_i are linearly independent so that σ is invertible. The Langevin SDE for $q, p \in \mathbb{R}^d$, the position and momenta of a particle of unit mass, is then

$$dq = p\,dt, \tag{2.5.1}$$

$$dp = -\gamma p\,dt - \nabla F(q)\,dt + \sigma\,dW. \tag{2.5.2}$$

Here $\gamma > 0$ to ensure a damped-driven Hamiltonian. In the case $d = 1$ and $\sigma = \sqrt{2}$, for example, there is a known invariant measure with density

$$\rho(p, q) \propto \exp\left\{-\gamma\left[\frac{p^2}{2} + F(q)\right]\right\}.$$

We assume that (1.5.4) holds with (1.5.5) being a prototypical example. This problem may be viewed as an analogue of Example 2.2.2 in that the noise process dW/dt is white—uncorrelated in time.

The Langevin equation (2.5.1), (2.5.2) is ergodic: the assumptions of Theorem 2.3.9 can be established, using the Lyapunov function (1.5.7) and the random analogue of (1.5.8) to establish Assumption 2.3.6. Because noise is only present in the momentum equation (2.5.2) it is not immediate that the smoothness and reachability (Assumption 2.3.8) follow, and certain vector field commutators need to be checked. See [16] for details.

A simple calculation, indicating why ergodicity might hold, is as follows. By Lemma 2.1.2

$$\mathbb{E}H(q(t), p(t) = \mathbb{E}H(q(0), p(0)) + \mathbb{E}\int_0^t \left[\frac{\sigma^2}{2} - \gamma\|p(s)\|^2\right] ds,$$

and so

$$\limsup_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{E}\|p(s)\|^2\,ds \leqslant \frac{\sigma^2}{2\gamma}.$$

This shows that the *Césaro average* of $\mathbb{E}\|p(s)\|^2$ is uniformly bounded in time. In fact

$$\mathbb{E}\|p(t)\|^2 \to \frac{\sigma^2}{2\gamma}$$

by ergodicity [16]. The split-step backward Euler method is also ergodic because it preserves Lyapunov structure; see [16].

2.6 Case study 2: particles in a random velocity field

Random velocity fields are frequently used in turbulence modelling to give analytically tractable caricatures of complex flows. In particular they can be fitted to predictions about energy spectra. The idea was pioneered by Kraichnan [13] and a useful overview of the subject is given in Section 6 of [14]. We describe a model for the motion of particles in a random two-dimensional incompressible velocity field.

Let $x \in \mathbb{T}^2$ (in this section $\mathbb{T}^2$ denotes the two-dimensional unit torus) denote the particle position and $\dot{x} \in \mathbb{R}^2$ its velocity. We assume that

$$\tau\ddot{x} = v(x, t) - \dot{x}, \tag{2.6.1}$$

where $v = \nabla^\perp \psi$. Here ψ is the stream-function, $\nabla^\perp$ denotes skew-gradient and hence $\nabla \cdot v = 0$. Thus x denotes the position of a particle moving according to Stokes' law in a two-dimensional incompressible velocity field v. We assume that

$$\psi = \sum_{k \in \mathcal{K}} y_k \cos(k \cdot x) + z_k \sin(k \cdot x),$$

where

$$\mathcal{K} = \{2\pi(k_1, k_2), k_i \in \{1, \ldots, M\}\}$$

and

$$dy_k = -\alpha_k y_k \, dt + \sqrt{\lambda_k} \, dB_k^y,$$
$$dz_k = -\alpha_k z_k \, dt + \sqrt{\lambda_k} \, dB_k^z.$$

Here the families $\{B_k^y\}_{k \in \mathcal{K}}$, $\{B_k^z\}_{k \in \mathcal{K}}$ are mutually independent families of i.i.d. standard Brownian motions. We assume that

$$\alpha := \min_{k \in cK} \{\alpha_k\} > 0.$$

The stream-function ψ may be viewed as the solution of a stochastic PDE of Ornstein-Uhlenbeck type, leading to the equation

$$d\psi + A\psi \, dt = dW,$$

where $A = -\Delta$ subject to periodic boundary conditions on $\mathbb{T}^2$, and W is a Wiener process in $L^2(\mathbb{T}^2)$ with covariance operator Q. With this notation the α_k are the eigenvalues of A whilst the λ_k are the eigenvalues of Q. We have taken M finite but infinite M can be handled.

We view ψ as the noise process in $C(\mathbb{R}, C^2(\mathbb{T}^2))$, driving the random dynamical system for particle motions in $\mathbb{T}^2 \times \mathbb{R}^2$. In this sense the problem is an analogue of Example 2.2.3 since the driving noise is coloured—correlated in time. See [23] for details of conditions on Q leading to sufficient regularity of ψ to obtain existence and uniqueness for (2.6.1).

Fig. 5 shows a typical velocity field at a fixed instant of time. (This should be compared with Fig. 3 shown earlier.) Fig. 6 shows the distribution of particles at time $t = 1$, starting from an initial configuration in which the particles are at rest and uniformly distributed on a grid.

This problem has a random attractor (see [23]), and the absorbing set calculation is very similar to that presented in the deterministic case, where $v(x, t) = v(x)$. The problem is also ergodic; again Theorem 2.3.9 can be used to establish this (see [17]).

Acknowledgments The authors are grateful to James Robinson, Hersir Sigurgeirsson, John Terry, Roger Tribe, Paul Tupper and Jon Warren for help in preparation of these notes. Hersir Sigurgeirsson created all the graphics related to Case Study 2. We are grateful to Anne Bourlioux, Martin Gander and Gert Sabidussi for giving us the opportunity, and necessary support, to enable us to give lectures based on these notes.

ARH has received support from the Engineering and Physical Sciences Research Council of the UK under grant GR/M06925, the Leverhulme Trust of the UK and the Fields Institute of Canada. AMS is supported by the Engineering and Physical Sciences Research Council of the UK under grant GR/N00340.

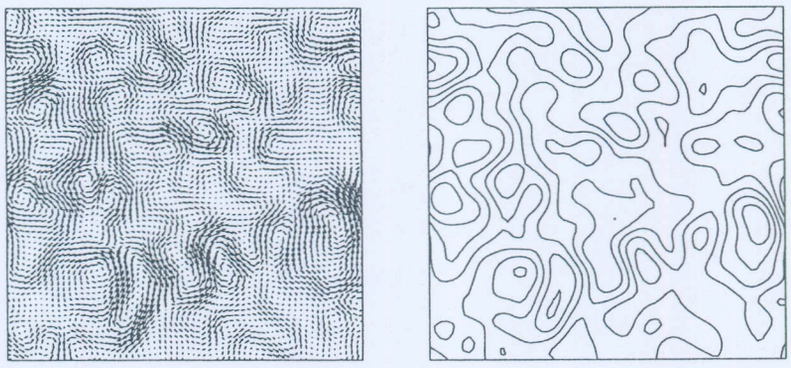


Figure 5: Random velocity field: (i) direction field (ii) streamlines.

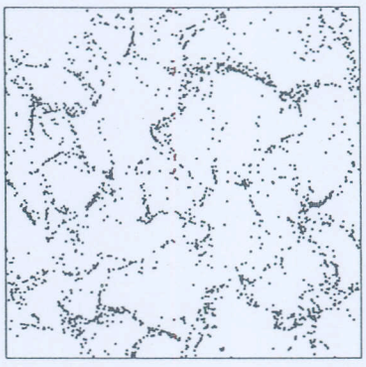


Figure 6: Particle distribution for (2.6.1).

References

[1] L. Arnold, *Random Dynamical Systems*, Springer-Verlag, New York, 1998.

[2] S. S. Artemiev and T. A. Averina, *Numerical Analysis of Systems of Ordinary and Stochastic Differential Equation*, VSP Press, Utrecht, 1997.

[3] D. Cheban, P. E. Kloeden and B. Schmalfuß, Pullback attractors in dissipative nonautonomous differential equations under discretization, *J. Dynam. Differential Equations* **13** (2001), 185–213.

[4] H. Crauel and F. Flandoli, Attractors for random dynamical systems, *Probab. Theory Related Fields* **100** (1994), 365–393.

[5] G. W. Ford and M. Kac, On the quantum Langevin equation, *J. Statist. Phys.* **46** (1987), 803–810.

[6] J. K. Hale, *Asymptotic Behaviour of Dissipative Systems*, Math. Surveys Monogr. **25** Amer. Math. Soc., Providence, RI, 1988.

[7] P. Hartman, *Ordinary Differential Equations*, Academic Press, New York, 1978.

[8] R. Z. Has'minskii, *Stochastic Stability of Differential Equations*, Sijthoff and Noordhoff, 1980.

[9] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, 1991.

[10] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, Academic Press, New York, 1975.

[11] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Appl. Math. **23**, Springer-Verlag, New York, 1992.

[12] P. Kloeden and B. Schmalfuß, Nonautonomous systems, cocycle attractors and variable time-step discretization, *Numer. Algorithms* **14** (1997), 141–152.

[13] R. H. Kraichnan, Diffusion by a random velocity field, *Phys. Fluids* **13** (1970), 22–31.

[14] A. Majda and P. R. Kramer, Simplified models for turbulent diffusion: theory, numerical modelling and physical phenomena, *Phys. Rep.* **314** (1999), 237–574.

[15] X. Mao, *Stochastic Differential Equations and Applications*, Horwood, Chichester, 1997.

[16] J. Mattingly, D. J. Higham, and A. M. Stuart, Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise, preprint, 2001.

[17] J. Mattingly and A. M. Stuart, Geometric ergodicity of some hypo-elliptic diffusions for particle motions, to appear in *Markov Process. Related Fields* (2002).

[18] S. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, Springer-Verlag, New York, 1992.

[19] B. Øksendal, *Stochastic Differential Equations*, Springer-Verlag, Berlin, 1998.

[20] J. C. Robinson, Stability of random attractors under perturbation and approximation, submitted to *J. Differential Equations* (2001).

[21] J. M. Sanz-Serna and M. P. Calvo, *Numerical Hamiltonian Problems*, Chapman & Hall, 1994.

[22] T. Shardlow and A. M. Stuart, A perturbation theory for ergodic Markov chains with application to numerical approximation, *SIAM J. Numer. Anal.* **37** (2000), 1120–1137.

[23] H. Sigurgeirsson and A. M. Stuart, Inertial particles in a random velocity field, submitted to *Stochastics Dynamics* (2002).

[24] A. M. Stuart and A. R. Humphries, *Dynamical Systems and Numerical Analysis*, Cambridge University Press, Cambridge, 1996.

[25] D. Talay, Second-order discretization schemes for stochastic differential systems for the computation of the invariant law, *Stochastics Stochastics Rep.* **29** (1990), 13–36.

[26] D. Talay, Approximation of upper Lyapunov exponents of bilinear stochastic differential systems, *SIAM J. Numer. Anal.* **28** (1991), 1141–1164.

[27] D. Talay, Approximation of the invariant probability measure of stochastic Hamiltonian dissipative systems with non globally Lipschitz coefficients, in: *Progress in Stochastic Structural Dynamics* (R. Bouc and C. Soize, eds.), Publ. Laboratoire Math. Appl. CNRS **152**, 1999.

[28] R. Temam, *Infinite Dimensional Dynamical Systems in Mechanics and Physics*, Appl. Math. Sci. **68**, Springer-Verlag, New York, 1989.

[29] G. E. Uhlenbeck, G. W. Ford, and E. W. Montroll, *Lectures in Statistical Mechanics*, Lectures in Appl. Math. **1**, Amer. Math. Soc., Providence, RI, 1963.