A Weak Bifurcation Theory for Discrete Time Stochastic Dynamical Systems

Cees Diks
Florian Wagener

CeNDEF, University of Amsterdam.

A weak bifurcation theory for discrete time stochastic dynamical systems

C. Diks and F.O.O. Wagener [*]

9th May 2006

Abstract

This article presents a bifurcation theory of smooth stochastic dynamical systems that are governed by everywhere positive transition densities. The local dependence structure of the unique strictly stationary evolution of such a system can be expressed by the ratio of joint and marginal probability densities; this 'dependence ratio' is a geometric invariant of the system. By introducing a weak equivalence notion of these dependence ratios, we arrive at a bifurcation theory for which in the compact case, the set of stable (non-bifurcating) systems is open and dense. The theory is illustrated with some simple examples.

1   Introduction

Bifurcation theory has been an extremely successful tool to investigate the qualitative properties of deterministic dynamical systems. Motivated by its success, there have been several attempts to develop bifurcation theory for stochastic dynamical systems, usually based on a classification of the shape of invariant densities of such systems. However, these attempts have run into the problem, already pointed out by Zeeman [15], that the shape of a probability density is not invariant under coordinate transformations.

For one-dimensional continuous time diffusions, a classification that is invariant under transformations has been proposed by Hartelman *et al.* [11, 14]. Inspired by their approach, we propose in this paper a classification for (strictly) stationary stochastic processes that are governed by smooth everywhere positive transition density functions. Let $\{X_t\}_{t \in \mathbb{Z}}$ be such a process, with probability density p^t, joint probability density of $X_{t_1}, \cdots, X_{t_n}$ denoted by $p^{t_1, \cdots, t_n}$, and transition

---

[*]Center for Nonlinear Dynamics in Economics and Finance (CeNDEF), Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, tel: +31 20 525 4034, fax: +31 20 525 4349, *C.G.H.Diks@uva.nl*, *F.O.O.Wagener@uva.nl*

probability density $\tau(x_{t+1}|x_t)$. Then $p^{t_1+h,\cdots,t_n+h} = p^{t_1,\cdots,t_n}$, because of strict stationarity, and

$$p^{t+1}(x_{t+1}) = \int \tau(x_{t+1}|x_t)p^t(x_t)\,\mathrm{d}x_t.$$

In particular, as $p^{t+1} = p^t = p$, the invariant probability density p is seen to be the solution of the integral equation $p(x) = \int \tau(x|y)p(y)\,\mathrm{d}y$ under the condition that $\int p(y)\,\mathrm{d}y = 1$; consequently, p is as least as smooth as τ.

The joint density $p^{t,t+1}$ does not depend on t and it is therefore equal to $p^{1,2}$. Moreover, the joint measure $p^{1,2}(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$ is absolutely continuous with respect to the product measure $p(x_1)p(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$; therefore, by the Radon-Nikodym theorem, the following function is well-defined:

$$f(x_1, x_2) \overset{\text{def}}{=} \frac{p^{1,2}(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2}{p(x_1)p(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2} = \frac{p^{1,2}(x_1, x_2)}{p(x_1)p(x_2)} = \frac{\tau(x_2|x_1)}{p(x_2)}.$$

We call the function f the *dependence ratio* of the system. Note that f is identically 1 if X_t and X_{t+1} are independent; the difference $|f(x_1, x_2) - 1|$ can therefore be seen as a measure of the local dependence structure of the stochastic dynamical system.

By construction, a dependence ratio is a geometric invariant of the underlying system, and it is therefore a suitable quantity to be at the foundation of a stochastic bifurcation theory. Several other local dependence measures have recently been described in the statistical literature (see e.g. [7], [8], and [9]). These measures are localised versions of the Pearson correlation coefficient, and as such are motivated entirely differently than our dependence ratio. In particular they do not share the geometrical invariance property.

Our concept of 'stochastic bifurcation' will not be an analogue of the concept of 'topological bifurcation' of deterministic systems. Recall that two deterministic systems are said to be topologically equivalent if one is induced by a homeomorphism from the other. It is easy to see, and it will be shown below, that we cannot build a theory of stochastic bifurcation on the parallel notion of 'strong equivalence', which defines two stochastic systems to be equivalent if the dependence ratio of the first is induced from the dependence ratio of the second by a coordinate transformation. The equivalence notion that we shall adopt is more like the weaker 'local topological equivalence' of deterministic systems, where two systems are equivalent if for all equilibria neighbourhoods exist such that the systems are topologically equivalent on corresponding neighbourhoods. Our analogous notion, which we call 'ratio equivalence', calls two dependence ratios equivalent if there is a coordinate transformation that maps critical points of one ratio to the corresponding critical points of the other ratios. As in the case of local topological equivalence, we shall show that for ratio equivalence the union of the open equivalence classes is an open and dense set in the space of all dependence ratios.
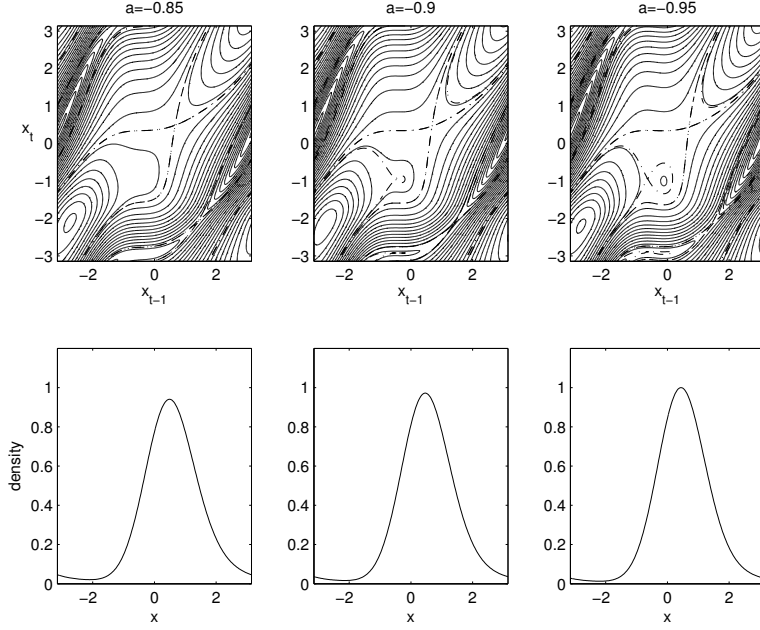
2

Figure 1: Level sets for the map $X_t = X_{t-1} + a \sin(X_{t-1}) + 0.25 \sin^2(X_{t-1}) + 0.25 + \varepsilon_{t+1}$ for decreasing values of a (top panels). The dashed-dotted lines, corresponding with the critical levels of the saddle points, are added for clarity. The lower panels show the invariant probability density of X_t.

We give figure 1 as an illustration. It shows the invariant probability densities and the dependence ratios of the stochastic dynamical system on the circle

$$X_t = X_{t-1} + a \sin(X_{t-1}) + 0.25 \sin^2(X_{t-1}) + 0.25 + \varepsilon_t \mod 2\pi$$

for several values of the parameter a; the ε_t are identically and independently distributed (IID) stochastic variables. We see that the number of critical points of the dependence ratio changes as a changes, and hence that the associated strictly stationary processes are qualitatively different. This change in the dynamics is not at all appearant from the invariant probabilities. Moreover, by a suitable change of coordinates in the system, the invariant probabilities could be made constant, whereas the transformed dependence ratio would still show the same bifurcation behaviour.

Plan of the article   In section 2, we quickly review the previously proposed notions of phenomenological and dynamical bifurcation of stochastic dynamical systems. We define a new equivalence relation, based on the dependence structure of the process in section 3. We show in particular that our equivalence relation has 'many' structurally stable elements and that it avoids some limitations of older notions. Finally, in section 4 we illustrate its usefulness by giving several applications. An appendix contains proofs.

3

2 Related literature

In this article, the term 'stochastic dynamical system' will always refer to a class of stochastic processes which are governed by a single transition probability density. An 'evolution' of such a stochastic dynamical system is a stochastic process of this class that is generated by an this transition probability density and an initial probability distribution.

2.1 Phenomenological bifurcations

The natural first attempt to attain at a classification of stochastic dynamical systems is to apply the Morse classification of real valued functions to invariant probability densities p of the system, see [3, 15]. The corresponding equivalence relation is that of smooth coordinate transformations of domain and range of p, the stable elements being Morse functions with all critical values distinct from each other. For the purposes of this article, we shall call the equivalence relation *P-equivalence*, in analogy with the associated bifurcation notion, which has been called *phenomenological bifurcation* or *P-bifurcation* (see Arnold [2], p. 471-473).

A serious limitation of the P–bifurcation, acknowledged in [15], is that the equivalence classes are not invariant under diffeomorphisms of the underlying space. For instance, let $\{X_t\}$ be a process on $\mathbb{R}^m$ with invariant density p_X, and let φ be an invertible transformation of $\mathbb{R}^m$. The density p_Y of $\{Y_t\}$, where $Y_t = \varphi(X_t)$, then is related to p_X through

$$p_X(x) = p_Y(\varphi(x)) \left|\det D\varphi(x)\right|.$$

We see that, in the language of physicists, the function value of the invariant density 'depends on the coordinates chosen'. Only if φ is volume-preserving, that is, if $\left|\det D\varphi(x)\right| = 1$ for all x, the form of the invariant density is guaranteed not to change. It is easy to construct examples on the real line where after a well-chosen transformation φ the densities p_X and p_Y are in different equivalence classes. Note that the volume-preserving diffeomorphisms comprise the class of Riemannian isometries that were proposed in [15] to be the admissible class of coordinate transformations.

Underlying this lack of invariance is the fact that a probability density $p(x)$, unlike the measure $p(x)\,\mathrm{d}x$, is not a geometrical invariant under general coordinate transformations. By consequence, P-equivalence is an inconvenient notion for practical applications: for instance, it might make a difference to the results whether data is recorded on a linear or a logarithmic scale.

As pointed out by Hartelman and co-workers [11, 14], it is possible to construct a 'coordinate-free' classification for one-dimensional continuous-time diffusions . By defining stochastic analogues of concepts used in catastrophe theory, they

arrived at a classification that is, unlike P-equivalence, invariant under monotonically increasing transformations of the real line, or more precisely, a classification that is invariant up to transformations homotopic to the identity mapping. This classification, together with invariant estimation considerations, leads to the empirical use of statistical quantities called level crossing statistics, which are invariant under monotonic transformations, that is, diffeomorphisms, of the real line.

Unfortunately level crossing statistics are not very natural in the discrete time setting, which is for instance the natural setting for many problems in economic dynamics. Although these statistics can be used for discrete time systems in principle, the corresponding classification would be rather restrictive, since discrete time dynamical systems are 'essentially richer' than discretely sampled continuous time diffusions, mainly because finite time transition densities induced by diffusions only represent a subclass of transition densities for discrete time dynamical systems.

2.2 Dynamical bifurcations

A second bifurcation notion for stochastic dynamical systems has been introduced by Ludwig Arnold and his co-workers (see [2] for an extensive exposition). We shall try to sketch this approach using a process $\{X_t\}$ on $\mathbb{R}$ of the form

$$X_{t+1} = g(X_t, \varepsilon_t), \tag{1}$$

with deterministic initial condition $X_0 = x_0 \in \mathbb{R}$, and where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables. The main idea is to consider this process as a deterministic dynamical system on an infinite dimensional phase space $\Omega \times \mathbb{R}$. The elements of Ω are the possible realisations $\omega = (\varepsilon_0, \varepsilon_1, \cdots)$ of the process $\{\varepsilon_t\}$. Introducing the projection $\pi(\omega) = \varepsilon_0$ and the shift $\sigma(\omega) = (\varepsilon_1, \varepsilon_2, \cdots)$, we have for instance that $\varepsilon_t = \pi \circ \sigma^t(\omega)$. Define now the map Φ on $\Omega \times \mathbb{R}$ by

$$\Phi(\omega, x) = (\varphi_1(\omega), \varphi_2(\omega, x)) = \big(\sigma(\omega), g\big(x, \pi(\omega)\big)\big).$$

This is a deterministic system; the stochastics are 'hidden' in the fact that the initial condition $\omega \in \Omega$ is unknown. The realisations X_t of the process (1) are the values of the second component of $\Phi^t(\omega, x)$. The map Φ is then called a random dynamical system on the phase space $\mathbb{R}$.

Note that Φ is a skew system: the shift dynamics φ_1 in the space Ω are driving the dynamics φ_2 in $\mathbb{R}$. For Φ, a *random fixed point* is defined as a map $\xi : \Omega \to \mathbb{R}$ that satisfies the invariance condition

$$\varphi_2(\omega, \xi(\omega)) = \xi(\varphi_1(\omega))$$

for all (or almost all) ω. Stability is now defined in the usual way: a random fixed point ξ is stable if all nearby orbits converge to ξ. Note that a stochastic dynamical system that is at a stable random fixed point has an invariant measure, which describes exactly the distribution of the random fixed point. So-called 'random' bifurcations are now defined as 'ordinary' bifurcations of the deterministic dynamical system Φ. For instance, a random, or, following the terminology in [2], *dynamical bifurcation* or *D-bifurcation* of a process occurs if a random fixed point loses stability.

At this point, a drawback of the notion of dynamical bifurcation becomes apparent: to determine stability of a random fixed point, two orbits of Φ with identical noise realisations have to be compared. This seems to make it rather difficult to apply the notion of D-bifurcation to practical problems (but see [4] and related literature).

The theory leading to the D-bifurcation remains close to traditional bifurcation theory for deterministic dynamical systems, in that it aims at characterising the full dynamical system rather than just invariant measures. However, as mentioned above, without further knowledge of the dynamics it is empirically impossible to distinguish between different dynamical systems producing time series with identical invariant measures. To avoid such differences at the level of the underlying dynamics that are undetectable empirically, we choose to leave aside this theory, and try rather to improve on the notion of P–equivalence.

3 Equivalence notions

In this section we introduce and motivate our bifurcation theory of of stochastic dynamical systems and we give some of its fundamental properties.

3.1 Definitions and assumptions

We want our theory to be applicable to stochastic dynamical systems of the form

$$X_{t+1} = g(X_t, \varepsilon_t), \tag{2}$$

where $g : \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}^m$ is such that $g(X_t, \cdot) : \mathbb{R}^k \to \mathbb{R}^m$ is a submersion, that is $\operatorname{ran} D_\varepsilon g \geq m$, and where ε_t are IID random variables taking values in $\mathbb{R}^k$. If g is smooth, and if the ε_t are distributed according to a smooth probability density function, then to this stochastic dynamical system there is associated a smooth transition probability density.

Taking a slightly more general setup, we consider a state space M that is either an open simply connected subset of $\mathbb{R}^m$ with smooth boundary, or the (compact) closure of a bounded open simply connected subset of $\mathbb{R}^m$ with smooth boundary, or a compact closed m-dimensional manifold. Motivated by the system

given in equation (2), we define a 'smooth' stochastic dynamical system as a smooth transition probability density $\tau : M \times M \to \mathbb{R}$: if the state X_t of the system at time t is distributed according to $p^t(x_t)\,\mathrm{d}x_t$, then X_{t+1} is distributed according to $p^{t+1}(x_{t+1})\,\mathrm{d}x_{t+1}$, where

$$p^{t+1}(x_{t+1}) = \int_M \tau(x_{t+1}|x_t)p^t(x_t)\,\mathrm{d}x_t. \tag{3}$$

Note that $\{X_t\}$ is a Markov process.

Assumption 1 *We shall assume throughout the paper that $\tau > 0$ on $M \times M$.*

This assumption implies that the integral equation

$$p(x) = \int_M \tau(x|y)p(y)\,\mathrm{d}y$$

has a unique solution $p : M \to \mathbb{R}$ satisfying $\int_M p(x)\,\mathrm{d}x = 1$ and $p(x) > 0$ for all $x \in M$. The function p is then the unique invariant probability density of the system; consequently, if X_0 is distributed according to $p(x)\,\mathrm{d}x$, then $\{X_t\}$ is a strictly stationary Markov process, and the X_t are identically distributed random variables. We denote the joint probability density of the random variables $(X_{t_1}, \cdots, X_{t_n})$ by $p^{t_1,\cdots,t_n}$. For instance, for a strictly stationary process we have $p^{t,t+1}(x, y) = p^{1,2}(x, y)$. To a strictly stationary first order Markov process with smooth invariant densities, we associate the *dependence ratio*

$$f(x_1, x_2) = \frac{p^{1,2}(x_1, x_2)}{p(x_1)p(x_2)} = \frac{\tau(x_2|x_1)}{p(x_2)}. \tag{4}$$

We make two remarks. First, if the dependence ratio f and the invariant density p are specified, the transition probability density τ is determined by equation (4). Second, as τ is smooth and $\tau > 0$ on $M \times M$, the associated dependence ratio f is smooth and takes values in $(0, \infty)$.

Assumption 2 *We shall assume that if M is compact, the dependence ratio f of a stochastic dynamical systems has no critical points on the boundary of $M \times M$.*

We note that dependence ratios are connected to copulas: in the simplest, two-dimensional case, a copula $C : [0, 1]^2 \to [0, 1]$ is a bivariate distribution function whose margins are uniformly distributed on $[0, 1]$. They arise as follows: let $\{X_t\}$ be a strictly stationary real-valued process with invariant probability density p, and let

$$F(x) = \int_{-\infty}^x p(y)\,\mathrm{d}y \quad \text{and} \quad F^{1,2}(x_1, x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} p^{1,2}(y_1, y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2$$

7

be its marginal and joint distribution functions. Then the copula $C(u_1, u_2)$ of this process is determined by the equation

$$F^{1,2}(x_1, x_2) = C(F(x_1), F(x_2)).$$

Now, differentiation of this relation with respect to x_1 and x_2 yields that

$$p^{1,2}(x_1, x_2) = \frac{\partial^2 C}{\partial u_1 \partial u_2}\big(F(x_1), F(x_2)\big)p(x_1)p(x_2),$$

and we see that the dependence ratio f of the process satisfies

$$f(x_1, x_2) = \frac{\partial^2 C}{\partial u_1 \partial u_2}\big(F(x_1), F(x_2)\big).$$

This association of dependence ratios with copulas neatly illustrates that the dependence ratio describes the whole stochastic dynamical system, and not merely its strictly stationary evolution. In the case of a smooth stochastic dynamical system on the real line with invariant probability $p(x)$ and dependence ratio $f(x_1, x_2)$, this is seen most readily by performing the coordinate change $\Xi_t = F(X_t)$, where $F(x) = \int_{-\infty}^{x} p(y)\,\mathrm{d}y$; the strictly stationary evolution of the transformed system is then uniformly distributed on the interval $[0, 1]$, and its transition probability density is then exactly equal to the transformed dependence ratio:

$$\tau(\xi_2|\xi_1) = f\left(F^{-1}(\xi_1), F^{-1}(\xi_2)\right).$$

We conclude that smooth stochastic dynamical systems on the real line that have strictly stationary evolutions are fully characterised, up to diffeomorphism, by their dependence ratios.

3.2 Structural stability and bifurcations

We recall briefly the fundamentals of bifurcation theory. The two main ingredients of any such theory are a topological space X and an equivalence relation between elements of X. An element f of X is *structurally stable* if there is a neighbourhood $N(f)$ such that all elements g in that neighbourhood are equivalent to f; that is $g \sim f$ for all $g \in N(f)$. Intuitively speaking, a structurally stable element f can be 'perturbed' slightly without being pushed out of its equivalence class. Such an element is sometimes called 'persistent'. Clearly, the equivalence class of any structurally stable element is an open set. A structurally stable equivalence class can be thought of as defining a set of elements of the same 'shape' or 'form' (see [13]): form remains 'stable' if perturbed slightly.

All elements of X that are not structurally stable are called *bifurcating*. This notion is usually familiar from the context of parametrised families: if λ is

8

some q-dimensional parameter, and $\lambda \mapsto f_\lambda$ a family of elements of X, then $\lambda = \lambda_0$ is a bifurcating parameter value of the family if f_{λ_0} is not structurally stable; it might be said that at bifurcating parameter values the 'form' of f_λ changes. Since the set of structurally stable elements is open, the set of bifurcating elements, and therefore also the set of bifurcating parameter values in a parametrised family, is closed.

An equivalence relation will give rise to a meaningful bifurcation theory on X only if there exist structurally stable elements at all. The most useful situation is attained if the set of structurally stable elements, while not consisting of a single equivalence class, is 'topologically big', since then we will be able to associate to 'most' elements a form. In a topological space, a set is 'big' if it is open and dense, or if it is at least a countable intersection of open and dense sets (a so-called 'generic' or 'second category' set, see [10]).

3.3 Strong equivalence

In the following, smooth stochastic dynamical systems are analysed and classified in terms of their unique strictly stationary evolutions.

A natural requirement to impose on an equivalence relation of smooth stochastic dynamical systems on M is that systems which only differ by a diffeomorphism of M, that is, which are the 'same' up to a coordinate change, fall in the same equivalence class. Let for instance $\{X_t\}$, $\{Y_t\}$ denote two stationary evolutions for which

$$Y_t = \varphi(X_t), \qquad \text{for all } t,$$

for some diffeomorphism $\varphi : M \to M$. We call the associated systems *strongly equivalent*. Denote the probability densities of X and Y respectively by an index X or Y. Since

$$
\begin{aligned}
p_X^{t_1,\cdots,t_n}&(x_{t_1}, \cdots, x_{t_n}) \\
&= p_Y^{t_1,\cdots,t_n}(\varphi(x_{t_1}), \cdots, \varphi(x_{t_n}))|\det D\varphi(x_{t_1})| \cdot \ldots \cdot |\det D\varphi(x_{t_n})|,
\end{aligned}
$$

it follows that

$$
\begin{aligned}
f_X(x_1, x_2) &= \frac{p_X^{1,2}(x_1, x_2)}{p_X(x_1) \cdot p_X(x_2)} \\
&= \frac{p_Y^{1,2}(\varphi(x_1), \varphi(x_2))|\det D\varphi(x_1)\det D\varphi(x_2)|}{p_Y(\varphi(x_1))|\det D\varphi(x_1)| \cdot p_Y(\varphi(x_2))|\det D\varphi(x_2)|)} \\
&= f_Y(\varphi(x_1), \varphi(x_2)).
\end{aligned}
\tag{5}
$$

We conclude that strongly equivalent systems have dependence ratios that are equal up to diffeomorphism.

If we took strong equivalence as the equivalence relation defining our bifurcation theory, in general we would obtain an uncountable infinity of equivalence

classes, and no class would be a neighbourhood to any of its points, that is, no system would be structurally stable and every system would be bifurcating. To see this in a simple example, assume that f_X and f_Y are two dependence ratios defined on the square $(-1,1) \times (-1,1) \subset \mathbb{R}^2$, and that they are given as

$$f_X(x_1, x_2) = \frac{2-\mu}{3} + x_1^2 + \mu x_2^2, \qquad f_Y(x_1, x_2) = \frac{2-\nu}{3} + x_1^2 + \nu x_2^2.$$

Taking the invariant density in both cases to be $p(x) = \frac{1}{2} I_{[-1,1]}(x)$, where $I_A(x)$ denotes the indicator function, we have specified two stochastic dynamical systems. The point $(0,0)$ is the only non-degenerate critical point for both f_X and f_Y; therefore, if f_X and f_Y are strongly equivalent, we should have that $\Phi(x_1, x_2) = (\varphi(x_1), \varphi(x_2))$ satisfies $\Phi(0,0) = (0,0)$. But there is not even a homeomorphism, much less a real-valued smooth diffeomorphism φ such that (5) holds simultaneously with $\varphi(0) = 0$, for the values of f_X and f_Y at $(0,0)$ are different if $\mu \neq \nu$. We see that every value of μ defines a different equivalence class.

3.4 Ratio equivalence

As we have seen, there are 'too many' equivalence classes if we take strong equivalence as our equivalence relation; put differently, equality of stationary evolutions up to diffeomorphism yields an equivalence relation that is too fine-grained to be useful. A coarser classification is obtained by retaining only certain topological information of the dependence ratio of the process. As this ratio is a geometric invariant, the classification will still be invariant under diffeomorphisms.

Recall the Morse classification of functions: two functions are of different shape if they have a different number of nondegenerate critical points. The number of such points is a numerical characteristic of the 'shape' of a function, and in fact we can build a classification where two functions are defined to be equivalent if they have the same number of nondegenerate critical points. Now, if we choose a suitable topology on the set of functions, we find that the equivalence classes are open sets, and that its members are structurally stable. Of course, usually we want also to include information about the type of critical point (Morse index).

The equivalence relation that we shall introduce on the space of smooth stochastic dynamical systems is based on the Morse (left-right) equivalence of functions, applied to the dependence ratio of the system. Notice however that, since dependence ratios are subject to certain restrictions, not every positive function on $M \times M$ is a dependence ratio, we have less freedom to perturb a given ratio, and therefore the properties of the Morse classification do not carry over automatically.

3.4.1 Topology

We need a function topology on the space of smooth stochastic dynamical systems and the space of their dependence ratios; we choose the C^2-topology, which is the

'coarsest' topology for which the number of nondegenerate critical points defines open equivalence classes. Recall that for functions $f : M \to \mathbb{R}$ defined on a compact manifold, an ε-neighbourhood $N_\varepsilon(f)$ of f in the C^2-topology consists of all functions g such that, with respect to a fixed Riemannian metric and the induced norms on the appropriate vector bundles $(T_x^* M, \cdots)$, all of them denoted by $\|\cdot\|$, we have

$$|f(x) - g(x)|, \ \|Df(x) - Dg(x)\|, \ \|D^2 f(x) - D^2 g(x)\| < \varepsilon,$$

for all $x \in M$. If M is a compact manifold-with-boundary, the inequalities are required to hold for all $x \in \mathrm{int}(M)$, in order that the derivatives remain well-defined. If M is a non-compact manifold, the constants $\varepsilon > 0$ are replaced by positive functions $\varepsilon(x) > 0$ on M in the above definition; the topology obtained is called the 'strong' C^2-topology (see e.g. [6]).

We use these topologies for the spaces of smooth stochastic dynamical systems, or rather the space of their transition probability densities $\tau : M \times M \to \mathbb{R}$, and the space of their dependence ratios. As explained in subsection 3.2, by specifying an equivalence notion on the topologised space of dependence ratios, we shall obtain a notion of structural stability and bifurcation of the associated stochastic dynamical systems that is invariant under coordinate changes.

3.4.2 Regular dependence ratios

A first rough formulation of our equivalence would be the following: we propose to call two stochastic dynamical systems equivalent, if every non-degenerate critical point of a certain type of the dependence ratio of the first system can be mapped to a critical point of the same type of the second system by a transformation of $M \times M$ that is induced by a diffeomorphism of M. We shall make this more precise, after having introduced some definitions.

Let $M^2 = M \times M$ be the Cartesian product of M with itself; denote by $\pi_\ell : M^2 \to M$, for $\ell = 1, 2$, the projection on the ℓ'th component

$$\pi_\ell(x_1, x_2) = x_\ell.$$

Recall the following definitions (see e.g. [5], subsections 10.2 and 10.4, p. 79 and p. 86 respectively). If $f : U \to \mathbb{R}$ is a twice continuously differentiable function defined on an open set $U \subset \mathbb{R}^n$, a point $x \in U$ is a *critical point* of f if the derivative of f vanishes at x: $Df(x) = 0$. The value $f(x)$ of f at a critical point x is called the *critical value* of f at x. A critical point x is *non-degenerate* if the Hessian matrix $Hf(x)$ corresponding to the second derivative $D^2 f(x)$ of f at x is invertible. The number of negative eigenvalues of this matrix is called the *(Morse) index* of the critical point. Clearly, the notions of critical point, critical value, index and non-degeneracy carry over to functions defined on manifolds.

Definition. *A twice differentiable dependence ratio $f : M^2 \to (0, \infty)$ is called* **regular** *if all its critical points are non-degenerate, if no two critical values are equal and if no two critical points have the same image under any projection π_ℓ, for $\ell = 1, 2$.*

3.4.3 Ratio equivalence on compact manifolds

In order to stay clear of topological complications, we only consider compact state spaces that have very simple boundaries.

Definition. *The set M is said to be a* **simple compact state space***, if either M is a closed compact manifold, or if M is the closure of an open bounded simply connected subset of $\mathbb{R}^m$ such that the boundary ∂M is a smooth manifold (in fact, a topological sphere).*

In this subsection, we shall always assume that M is a simple compact state space, unless mentioned otherwise. Note that the restrictions on M are imposed in order not to have to deal with topological side issues.

If M is a simple compact state space, a regular dependence ratio has only finitely many critical points $\xi_1, \cdots, \xi_k$; assumption 2 implies that none of these lie on the boundary of M^2. We assume that the points ξ_i are ordered such that the corresponding critical values $v_i = f(\xi_i)$ are in ascending order, that is, $v_i < v_j$ if $i < j$. We associate to the critical point ξ_i its index t_i (see subsection 3.3). Note that $0 \le t_i \le 2m$. In this way we obtain the *index sequence* $t(f) = (t_1, \cdots, t_k)$ of a regular dependence ratio f.

Definition. *Assume that M is a simple compact state space; let two smooth stochastic dynamical systems on M be given with everywhere positive transition probability and with dependence ratios $f, g : M^2 \to (0, \infty)$. These systems, and their associated dependence ratios, are said to be* **compact ratio equivalent***, if either both f and g are non-regular, or if f and g are both regular and*

1. *their index sequences are equal;*

2. *there is a diffeomorphism $\varphi : M \to M$, homotopic to the identity mapping on M, such that the* **induced diffeomorphism** *$\Phi : M^2 \to M^2$ defined as*

$$\Phi(x_1, x_2) = \Big(\varphi(x_1), \varphi(x_2) \Big) \tag{6}$$

   maps the i'th critical point of f to the i'th critical point of g.

It follows from the first point that the number of critical points of f and g is equal as well. We have the following proposition.

Proposition 1. *If M is a simple compact state space, then*

1. *a smooth stochastic dynamical system on M with everywhere positive transition density is structurally stable under compact ratio equivalence if and only if its dependence ratio is regular;*

2. *the set of stochastic dynamical systems on M that are structurally stable under compact ratio equivalence is open and everywhere dense in the C^2 topology.*

The proof of this proposition can be found in appendix A.

The proposition tell us that ratio equivalence has desirable properties, as we can characterise all structurally stable systems, and as these form an open and dense set in the space of all systems. In particular, it implies that we can build a bifurcation theory of stochastic dynamical systems on simple compact state spaces, based on the notion of compact ratio equivalence.

3.4.4   Ratio equivalence for non-compact manifolds

Though the results for the case of simple compact state spaces M are already useful in themselves, in practice most stochastic dynamical systems are defined on the non-compact manifold $\mathbb{R}^m$. In this section, we investigate the case that M is an open simply connected subset of $\mathbb{R}^m$, possibly $\mathbb{R}^m$ itself. The direct generalisation of the notion of compact ratio equivalence is given in the following definition.

Definition.   *Let M be an open simply connected subset of $\mathbb{R}^m$ such that its boundary ∂M is a smooth manifold. Two smooth stochastic dynamical systems, as well as their dependence ratios on $M \subset \mathbb{R}^m$, are **weakly ratio equivalent**, if there is a simple compact state space $K_0 \subset M$, such that for every simple compact state space $K \supset K_0$, the dependence ratios of the systems restricted to K are compact ratio equivalent on K^2.*

As the following example shows, this notion is unfortunately too weak for our purposes.

Example. Consider two stationary evolutions $\{X_t\}$ and $\{Y_t\}$ on the interval $M = (-1, 1)$ with invariant densities $p(x) = \frac{1}{2}I_{(-1,1)(x)}$ and dependence ratios

$$f_X(x_1, x_2) = 1 - \frac{1}{2}x_1 x_2 + \frac{1}{4}x_1^3, \qquad \text{and} \qquad f_Y(x_1, x_2) = 1 + \frac{1}{2}x_1 x_2 - \frac{1}{4}x_1^3.$$

Both ratios have a unique critical point of index 1 at the origin, and hence they are ratio equivalent on compact sets. But if we consider the values of f_X and f_Y along the curve $\gamma(t) = (t, t)$ as $t \uparrow 1$, we note that $f_X \circ \gamma(t) \to \inf_{M^2} f_X$, while $f_Y \circ \gamma(t) \to \sup_{M^2} f_Y$. Weak ratio equivalence is not sufficiently fine to distinguish between these systems. ∎

Definition.   *Let M be an open simply connected subset of $\mathbb{R}^m$ such that its boundary ∂M is a smooth manifold. If there exists a family $\{M_t\}$ of bounded open simply connected subsets of M, such that*

1. *the boundary ∂M_t is a smooth submanifold of M;*

2. *$M_t \subset M_{t'}$ if $t < t'$;*

3. *$\bigcup_t M_t = M$;*

*then we call $\{M_t\}$ an **exhaustion** of M.*

Note that the closure $\overline{M_t}$ of M_t is a simple compact state space.

A convenient way to define an exhaustion of M is to take a differentiable function $J : M \to \mathbb{R}$ with the following properties. We fix a point $x_0 \in M$ and we require that $J(x_0) = 0$, $J(x) > 0$ for all $x \neq x_0$, $J(x) \to \infty$ as $\|x - x_0\| \to \infty$ or $x \to \partial M$, and finally that J has no other critical points except x_0. Then $M_t = \{x \in M : J(x) < t\}$ is an exhaustion of M.

Consider the set

$$\partial M_t^2 \subset M^2.$$

This set can be decomposed into three component manifolds $C_t^1 = M_t \times \partial M_t$, $C_t^2 = \partial M_t \times M_t$ and $C_t^3 = \partial M_t \times \partial M_t$. In the important special case that M is one-dimensional, the component $C_t^3 = \partial M_t \times \partial M_t$ consists of four points. By definition, we consider these as non-degenerate critical points, associating the index 0 to them by default. We also define $C_t^0 = M_t \times M_t$.

In the following three definitions, M is a manifold with exhaustion $\{M_t\}$ and with decomposition $\{C_t^j\}_{j=0}^3$ of the closure of ∂M_t^2. Moreover, the restriction of f to C_t^j is denoted by f_t^j, for $j = 0, \cdots, 3$.

Definition.   *A smooth stochastic dynamical system with dependence ratio f on M^2 is **well-behaved at infinity** if there are constants $c_t, T > 0$ such that for every $t > T$ and every j:*

1. *if $\dim C_t^j > 0$, there is a compact set $K_t^j \subset C_t^j$ such that $\|Df_t^j(x)\| > c_t$ if $x \in C_t^j \backslash K_t^j$, and*

2. *f_t^j is weakly ratio equivalent to $f_{t'}^j$ on C^j for all $t, t' > T$.*

 This definition is illustrated in figure 2.

Definition.   *A smooth stochastic dynamical system with dependence ratio f on M^2 is **well-behaved** if f is well-behaved at infinity and f_t^j is regular on C_t^j for every j and every $t > T$.*
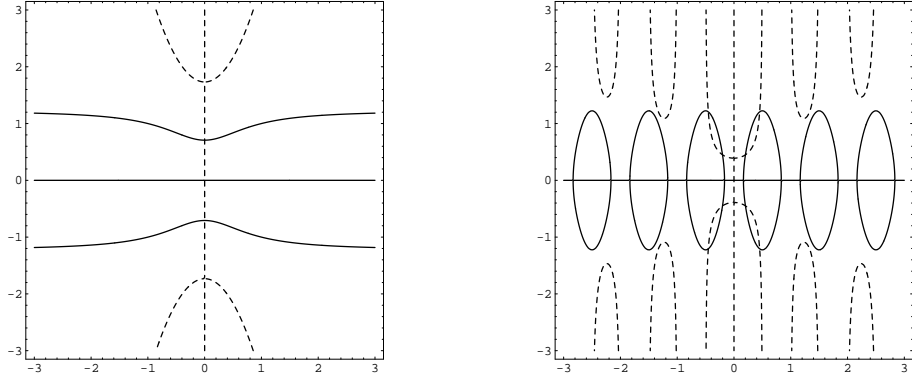
14

Figure 2: Well-behavedness at infinity. For dependence ratios $f : \mathbb{R}^2 \to \mathbb{R}$, the curves indicate the sets $\{\frac{\partial f}{\partial x} = 0\}$ (dashed) and $\{\frac{\partial f}{\partial y} = 0\}$ (solid). The dependence ratio on the left is well-behaved at infinity. For the dependence ratio on the right, note that f_n^2 (that is, f restricted to $C^2 = \{-n\} \times (-n, n) \cup \{n\} \times (-n, n)$) cannot be weakly ratio equivalent to $f_{n+\frac{1}{2}}^2$ for any integer n, as the number of critical points is different.

Definition. *If M is a manifold with exhaustion $\{M_t\}$, two well-behaved smooth stochastic dynamical systems with dependence ratios f and g are called* **ratio equivalent***, if there is a value of t such that f_t^j and g_t^j are weakly ratio equivalent for every j.*

Note that if f and g are weakly ratio equivalent on each component C_t^j for a single value $t > T$, they are in fact equivalent for all such values, since $f_t^j \sim f_{t'}^j$ for all $t, t' > T$.

Example. The two weakly equivalent ratios f_X and f_Y introduced at the end of the previous subsection are not ratio equivalent. Set $a_t = t/(t + 1)$, and consider the exhaustion $I_t = (-a_t, a_t)$ of $(-1, 1)$. Note that $\partial(I_t \times I_t)$ can be decomposed into

$$
\begin{aligned}
C_t^1 &= (-a_t, a_t) \times \{-a_t, a_t\}, \\
C_t^2 &= \{-a_t, a_t\} \times (-a_t, a_t), \\
C_t^3 &= \{-a_t, a_t\} \times \{-a_t, a_t\}.
\end{aligned}
$$

Restricted to C_t^1 and C_t^2, neither f_X nor f_Y have any critical points. The set C_t^3 consists of four isolated critical points, which are critical by definition. The maximum of f_X restricted to C_t^3 is assumed in the points $(a_t, -a_t)$, whereas f_Y takes its minimum there. Since the only diffeomorphism of C_t^3 homotopic to the identity is the identity itself, corresponding critical points of f_X and f_Y cannot be mapped onto each other. ∎

The following propositions describe the topological properties of ratio equivalence.

15

The results are weaker than in the compact case, as was to be expected; we obtain that well-behaved processes are stable elements of ratio equivalence. However, restricted to the space of processes that are well-behaved at infinity, the well-behaved processes form again an open and dense set.

Proposition 2. *On an open simply connected subset of $\mathbb{R}^m$, a well-behaved smooth stochastic dynamical system with everywhere positive transition probability density is stable with respect to the strong topology under ratio equivalence.*

The proof of this proposition is given in appendix A.

4 Examples

4.1 Stochastic dynamics on the circle

As an illustration of a stochastic dynamical system on a compact manifold, we consider the system on the unit circle $M = S^1$ defined by

$$X_{t+1} = X_t + a\sin(X_t) + 0.25\sin^2(X_t) + 0.25 + \varepsilon_{t+1} \mod 2\pi, \qquad (7)$$

with $\{\varepsilon_t\}$ IID and $N(0, \sigma^2)$ distributed. The state variable is taken modulo 2π; we represent states by points on the interval $[-\pi, \pi)$. For the above system we fix σ at the value 0.7 and consider qualitative changes in the stochastic dynamics as a varies. The term $0.25(\sin^2(X_{t-1}) + 1)$ is added to break the $x \mapsto -x$ symmetry of the dynamics. In the symmetric case some particular additional properties arise which will be discussed in the next subsection.

Figure 3 shows a contour plot for the dependence ratio $f(x_1, x_2)$ for values of a decreasing from -0.85 to -0.95. For $a = -0.85$, the contour plot shows two extrema, a maximum and a minimum, together with two saddle points. These are the minimal number of critical points of each type that can be attained for a nondegenerate function f on the torus $M_2 = S^1 \times S^1$. As the bifurcation parameter a decreases, the system shows a stochastic bifurcation. Between $a = -0.85$ and $a = -0.9$ the dependence ratio develops a new saddle point and a new local extremum near $(x_{t-1}, x_t) = (0, -1)$. Upon decreasing a further, another bifurcation takes place, where a new saddle point and a local minimum (near $(x_{t-1}, x_t) = (0, -3)$) appear.

4.2 Stochastic dynamics on the real line

As an example of dynamics on $\mathbb{R}$ we consider

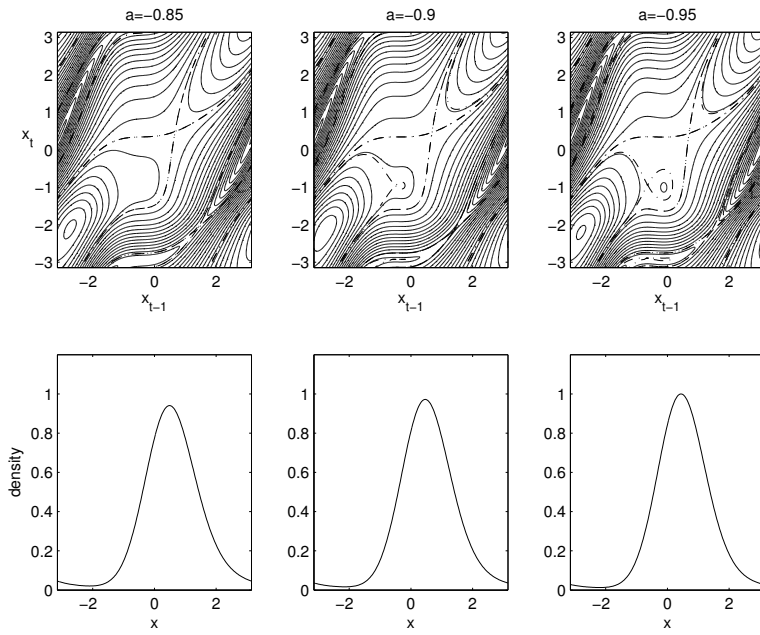$$X_{t+1} = \tanh(aX_t) + \varepsilon_{t+1}. \qquad (8)$$

Figure 3: Level sets for the map $X_t = X_{t-1} + a\sin(X_{t-1}) + 0.25\sin^2(X_{t-1}) + 0.25 + \varepsilon_{t+1}$ for decreasing values of a (top panels). The dashed-dotted lines, corresponding with the critical levels of the saddle points, are added for clarity. The lower panels show the invariant probability density of X_t.
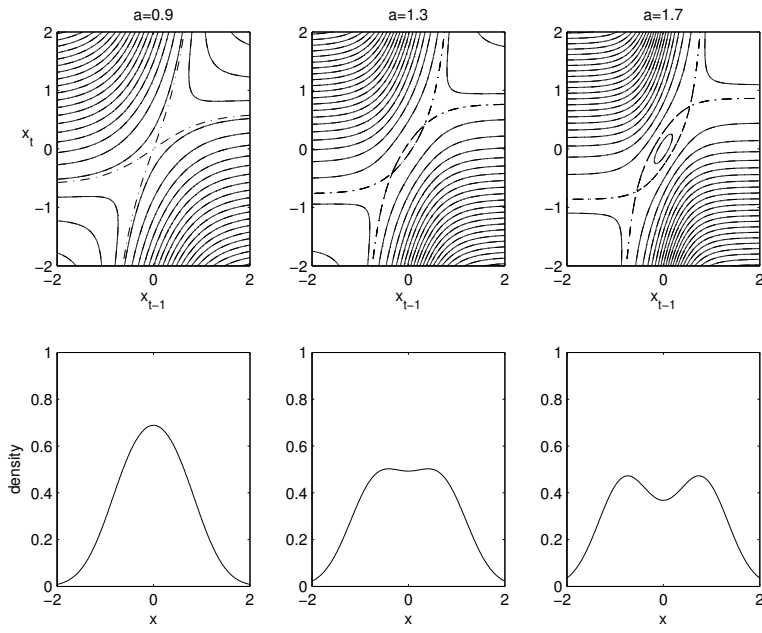
Figure 4: Level sets for the map $X_{t+1} = \tanh(aX_t) + \varepsilon_{t+1}$ with $\varepsilon_{t+1} \sim N(0, 0.5^2)$ for increasing values of a (top panels). The dashed-dotted lines correspond to the critical levels of the saddle points. The corresponding marginal density functions are shown in the lower panels.

Figure 4 shows the level sets of the dependence ratio and the corresponding invariant probability density function for this map with $N(0, \sigma^2)$ distributed noise, taking $\sigma = 0.7$.

Note that this example is special in that the dynamics is symmetric with respect to multiplication of reflecting the X-values in zero. Such symmetries are encountered often in families of models for real-world phenomena, in particular in cases where there are underlying physical reasons for assuming symmetry. In contrast with the previous example (the map on the circle) it can be observed in Fig. 3 that the invariant density become bimodal roughly when the dependence ratio changes qualitatively. Although beyond the scope of this paper, it can be shown analytically that the coincidence is perfect in this case. This is related to the symmetry of the map, together with the fact that the noise is additive and has a symmetric distribution, independent of the state X_t. In those cases the stochastic bifurcation based on dependence ratios coincides with a phenomenological bifurcation (P-bifurcation). It should be noted that this coincidence arises only in specific coordinates. As described above, a P-bifurcation can always be 'transformed away' by a change of coordinates.

It can be observed that the bifurcation parameter value differs from that of the analogous deterministic system ($\sigma = 0$): for the $\tanh$ map the stochastic analogue of the usual pitchfork bifurcation at $a = 1$ is shifted to a larger value of a. Apparently the value of the bifurcation depends on the noise level. A natural question, therefore, is whether for increasing noise levels the bifurcation parameter merely shifts, or whether the bifurcation can disappear altogether.

Intuitively, if the map is bounded and has a small range relative to the noise level, the dynamics is mainly governed by the noise and the deterministic part has little influence on the dynamics. In fact a simple argument shows that if the noise is fixed at a sufficiently large level, and if the family of odd maps $\{g_a\}$ is uniformly bounded, then there is no phenomenological bifurcation at $x = 0$, and therefore also no ratio bifurcation at $(x_1, x_2) = (0, 0)$, for symmetric processes of the form

$$X_{t+1} = g_a(X_t) + \varepsilon_{t+1}. \tag{9}$$

The argument runs as follows. By stationarity the invariant density p satisfies

$$p(x) = \int \frac{1}{\sigma} h\left(\frac{x - g_a(y)}{\sigma}\right) p(y) \, \mathrm{d}y,$$

where $h(\cdot)$ is the probability density function of the noise. A necessary condition for $p(x)$ to have a local minimum at $x = 0$ is that $p''(0) > 0$, where

$$p''(0) = \int \frac{1}{\sigma} h''\left(\frac{-g_a(y)}{\sigma}\right) p(y) \, \mathrm{d}y.$$

Since h is a unimodal probability density function, its second derivative $h''(x)$ is negative in a neighbourhood of $x = 0$. It follows that, for g_a uniformly bounded, for

large σ the integral on the right hand side of the last equation may remain negative as a varies.

4.3 Estimated dependence ratios from time series

In order to see whether dependence ratios can be used for classification of processes of which only a time series is available, a common situation in empirical applications, we estimate dependence ratios from simulated time series. We generate relatively short series $\{X_t\}$ from the stochastic models considered earlier in this section; we estimate from these series bivariate invariant densities and use them to reconstruct the dependence ratios. It is well known [1, 12] that fixed bandwidth nonparametric kernel density estimates become rather poor in regions with only few observations. One way to avoid this would be to use a data driven adaptive bandwidth which depends on the density locally, becoming larger as fewer observations are present locally. Instead of using an adaptive bandwidth we suggest, for real valued time series, to transform the data using the probability integral transform, that is, we construct

$$U_t = \widehat{F}_X(X_t) = \frac{\text{rank of } X_t \text{ among } \{X_s\}_{s=1}^{N}}{N}.$$

This amounts to transforming the invariant distribution to a uniform distribution on the unit interval, which tends to stabilise the estimation of the dependence ratio as the marginals no longer need to be estimated. The estimated empirical dependence ratio is then equal to the empirical copula density

$$\widehat{f}(u_1, u_2) = \frac{1}{N-1} \sum_{t=1}^{N-1} K_b(u_1 - U_i, u_2 - U_{i+1}).$$

Here $K_b(u_1, u_2)$ is a bivariate probability kernel, which we take to be the commonly used Gaussian kernel:

$$K_b(u_1, u_2) = \frac{1}{\sqrt{2\pi b}} \mathrm{e}^{-(u_1^2 + u_2^2)/(2b^2)}.$$

To avoid 'probability mass' from disappearing out of the unit square by this smoothing procedure, we impose periodic boundary conditions for $M = S^1$ and reflecting boundary conditions for $M = \mathbb{R}$.

Figure 5 shows level sets of the empirical dependence ratio obtained from time series of length 4000 from the symmetric hyperbolic tangent map given in equation (8) for different parameter values. The dependence ratio is estimated by smoothing the empirical copula with a bivariate normal probability density function (bandwidth $b = 0.07$). The empirical dependence ratio clearly reflects the fine structure of the theoretical dependence ratio. Figure 6 shows an attempt at performing a similar reconstruction for the asymmetric sine map given by equation (7). In
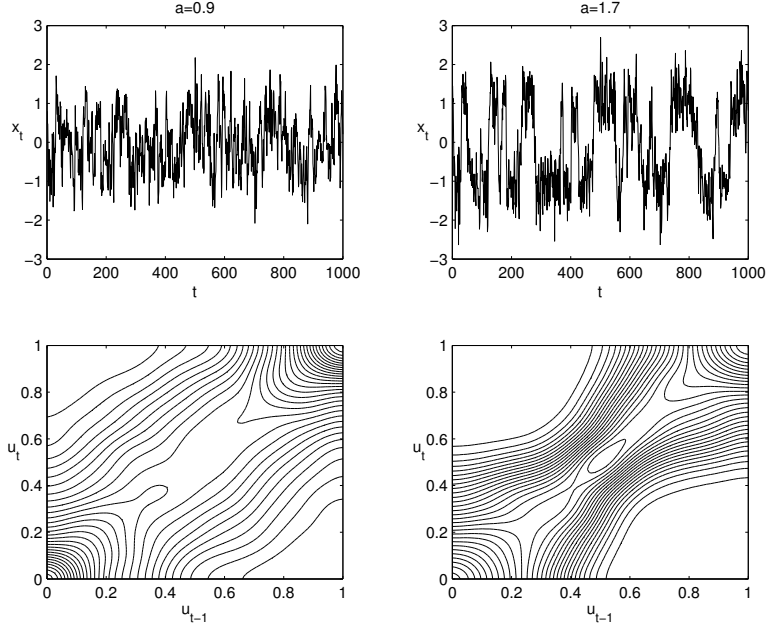
Figure 5: First 1000 values (top panels) of 4000 consecutive X_t-values generated by the map $X_{t+1} = \tanh(aX_t) + \varepsilon_{t+1}$ with $\varepsilon_{t+1} \sim N(0, 0.5^2)$. for $a = 0.9$ (left) and $a = 1.7$ (right). The lower panels show the corresponding empirical level sets estimated with a Gaussian kernel (bandwidth $b = 0.07$).

this case the topology of the reconstructed level sets does not correspond with that obtained earlier; this is due to estimation error. Probably longer time series (along with smaller bandwidths for the smoothers) are required for this case. We consider the optimal estimation and the related issue of data requirements for estimating dependence ratios as an important area for future research.

A   Proofs of the topological properties

In this appendix, the topological properties given in section 3 are proved.

A.1   Proofs of the propositions

We repeat the statement of proposition 1 for the convenience of the reader.

Proposition 1.   *If M is a simple compact state space, then*

1. *a smooth stochastic dynamical system on M with everywhere positive transition density is structurally stable under compact ratio equivalence if and only if its dependence ratio is regular;*
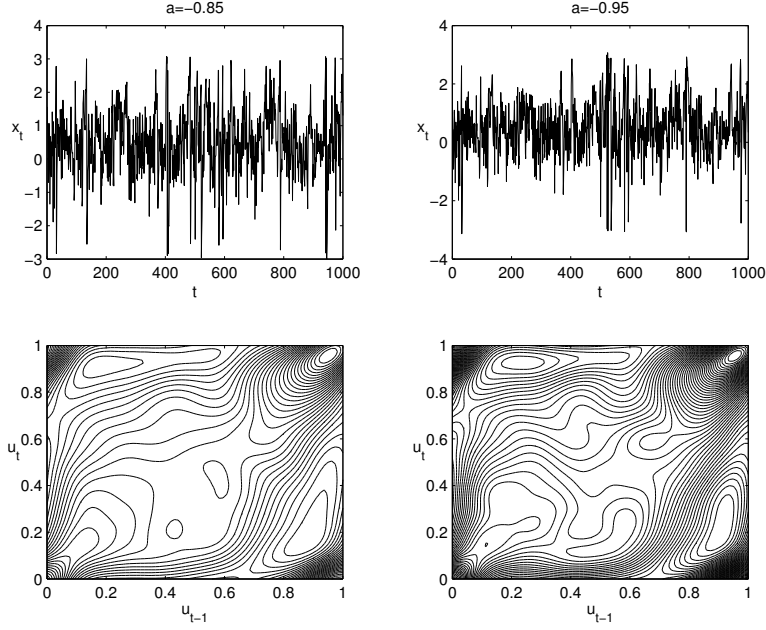
21

Figure 6: First 1000 values (top panels) of 4000 consecutive X_t-values generated by the map $X_{t+1} = X_t + a\sin(X_t) + 0.25\sin^2(X_t) + 0.25 + \varepsilon_{t+1}$ with $\varepsilon_{t+1} \sim N(0, 0.7^2)$ for $a = -0.85$ (left) and $a = -0.95$ (right). The lower panels show the level sets of the corresponding empirical dependence ratio.

2. *the set of stochastic dynamical systems on M that are structurally stable under compact ratio equivalence is open and everywhere dense in the C^2 topology.*

Proof.
This proposition is a direct corollary from the following two lemmas.

Lemma 1. *Let M be a simple compact state space. If $f : M^2 \to \mathbb{R}$ is a regular dependence ratio of a stochastic dynamical system, then there is a constant $\varepsilon > 0$ such that every dependence ratio $g \in N_\varepsilon(f)$ is regular and equivalent to f.*

Lemma 2. *Let M be a simple compact state space, and let a stochastic dynamical system be given with invariant probability density p and dependence ratio f. Then for every $\varepsilon > 0$ there is a second system with the same invariant probability density and regular dependence ratio g such that $g \in N_\varepsilon(f)$.*

We take for the moment these lemmas for granted. Then we infer from lemma 1 that regular ratios are structurally stable. Conversely, if f is a structurally stable ratio, there is a neighbourhood $N_\varepsilon(f)$ such that every $g \in N_\varepsilon(f)$ is equivalent to f. But as the regular ratios are dense, according to lemma 2, there is a system with the same invariant probability and with a regular dependence ratio $g \in N_\varepsilon(f)$ such

22

that g is equivalent to f. By definition of equivalence, the ratio f itself has to be regular. This finishes the proof of the first statement.

For the second statement, denote the transition probability density of the first system by τ_f and of the second system by τ_g. Note that since

$$\tau_f(x_2|x_1) = f(x_1, x_2)p(x_2), \qquad \tau_g(x_2|x_1) = g(x_1, x_2)p(x_2),$$

the fact that $g \in N_\varepsilon(f)$ implies that $\tau_g \in N_{\tilde{\varepsilon}}(\tau_f)$ and that $\tilde{\varepsilon} \to 0$ as $\varepsilon \to 0$. The proposition now follows from lemma 2. ∎

The following result is essentially a corollary of the results for simple compact state spaces.

Proposition 2. *Let M be an open simply connected subset of $\mathbb{R}^m$ such that the boundary ∂M is a smooth manifold. A well-behaved smooth stochastic dynamical system is structurally stable under ratio equivalence with respect to the strong topology.*

Proof.
Let f be the dependence ratio of the system, and let $\{M_t\}$ be an exhaustion of M. Since the system is well-behaved, there are $T > 0$ and $c_t > 0$ such that for every $j = 0, \cdots, 3$:

1. for every component $\{C_t^j\}$ of the closure of M_t^2 the restriction f_t^j of f to C_t^j is weakly ratio equivalent to $f_{t'}^j$, and

2. there is a compact set K_t^j such that $|Df_t^j(x)|_x > c_t$ if $x \in C_t^j \backslash K_t^j$.

Then there is for every $t > T$ and every j a constant $\varepsilon_t^j > 0$ such that for every $g \in N_{\varepsilon_t^j}(f)$ in the C^2-topology on M_t^2, the restriction g_t^j of g to C_t^j is weakly ratio equivalent to f_t^j. Let $\varepsilon_t = \min_j \varepsilon_t^j$, and $\varepsilon(x) = \max\{\varepsilon_t \,|\, x \in M_t^2\}$. It follows that $N_{\varepsilon(x)}(f)$ is an open neighbourhood of f in the strong C^2-topology, such that every $g \in N_{\varepsilon(x)}(f)$ is ratio equivalent to f. ∎

A.2 Proofs of the lemmas

It remains to demonstrate the lemmas. Note that they are similar to the analogous statements about regular functions in the Morse theory. The main technical difficulty is that the set of smooth dependence ratios on M^2 is a proper subset of the space of smooth functions on M^2, and therefore we have less freedom to construct perturbations of a function. This implies that while the proof of lemma 1 is a simple extension of the analogous arguments of Morse theory, the proof of lemma 2 is more involved.

We begin by recalling the following standard technical result, the proof of which is left to the reader. Note that we denote a ball of radius r around 0 by B_r, that is, $B_r = \{x \in \mathbb{R}^k \mid \|x\| < r\}$; also we introduce

$$\|f - g\|_{C^2(U)} = \max_{0 \leq j \leq 2} \max_{x \in U} |D^j f(x) - D^j g(x)|.$$

Lemma 3. *Let $U \subset \mathbb{R}^k$ be a bounded open set, and let $f : U \to \mathbb{R}$ be a C^2 function with $Df(0) = 0$ and $Hf(0)$ non-degenerate. Then there exist constants $\delta_0, \eta_0 > 0$ such that $B_{\delta_0} \subset U$ and that for every $0 < \delta \leq \delta_0$ and $0 < \eta \leq \eta_0$ there is an $\varepsilon > 0$, such that every function g satisfying $\|f - g\|_{C^2(U)} < \varepsilon$ has a unique non-degenerate critical point $\bar{y} \in B_\delta$ with $|g(\bar{y}) - f(0)| < \eta$, with g having the same index at $\bar{y}$ as f at 0.*

Proof of lemma 1. We give the proof for the case that M is the closure of a bounded open simply connected subset of $\mathbb{R}^m$. The generalisation to compact manifolds is straightforward but notationally awkward, and it is therefore left to the reader.

Let $x = (x_1, x_2)$ and $y = (y_1, y_2)$ denote points in M^2. Note that $\pi_\ell(x) = x_\ell$ etc. We shall first show that if $\varepsilon > 0$ is sufficiently small, then every $g \in N_\varepsilon(f)$ has the same index sequence as f; this part of the proof is standard. Afterwards, we shall construct a diffeomorphism on M, homotopic to the identity, which maps the critical points of g to the corresponding critical points of f.

Let $\xi_1, \cdots, \xi_k$ be the critical points of f, ordered such that $v_i = f(\xi_i) < f(\xi_j) = v_j$ if $i < j$. Put $v_0 = 0$; then $v_0 < v_1$. We set

$$\zeta = \min_{0 \leq i < j \leq k} |v_i - v_j|, \qquad \sigma = \min_{1 \leq i < j \leq k} \left\{ \|\pi_1(\xi_i) - \pi_1(\xi_j)\|, \|\pi_2(\xi_i) - \pi_2(\xi_j)\| \right\};$$

then ζ is the smallest absolute difference of two critical values, and σ is the smallest distance of two projections of critical points on M.

Choose $0 < \delta < \sigma$ such that the sets $U_i = \xi_i + B_\delta$ and $U_j = \xi_j + B_\delta$ are contained in M^2 and have empty intersection for any $i \neq j$; this is possible by assumption 2. Moreover, we set $f_i(x) = f(\xi_i + x)$ for $x \in B_\delta$. By assumption $Df_i(0)$ vanishes and $Hf_i(0)$ is nondegenerate. By lemma 3, we can find $\varepsilon > 0$, such that every function g_i defined on $\xi_i + B_\delta$ with $\|f_i - g_i\|_{C^2} < \varepsilon$ has a unique nondegenerate critical point y_i in B_δ, with $|f_i(0) - g_i(y_i)| < \zeta/2$ and with y_i having the same index t_i for g_i as 0 has for f_i.

Let $C = M^2 \setminus \bigcup_i U_i$; note that C is compact, and that $Df \neq 0$ on C. Therefore, if necessary by choosing $\varepsilon > 0$ smaller, we obtain that if $g \in N_\varepsilon(f)$, then $Dg \neq 0$ on C as well. This shows that the index sequences of f and g are equal.

We have to provide a diffeomorphism $\varphi : M \to M$, homotopic to the identity, such that

$$\Phi(x) = \big(\varphi \circ \pi_1(x), \varphi \circ \pi_2(x)\big)$$

24

maps the critical points $\xi_i + y_i$ of g to the critical points ξ_i of f. Note that by the choice of δ, no two projections of the sets U_i on M intersect:

$$\pi_1(U_i) \cap \pi_2(U_j) = \emptyset, \qquad \text{for all} \quad 1 \leq i, j \leq k.$$

For given i and ℓ, consider on $\pi_\ell(U_i)$ a differentiable curve $\gamma(t)$, defined for $0 \leq t \leq 1$, such that $\gamma(0) = \pi_\ell(\xi_i)$ and $\gamma(1) = \pi_\ell(\xi_i + y_i)$. Construct a vector field $X_{i\ell}$ on M such that $\dot{\gamma}(t) = X_{i\ell}(\gamma(t))$ for $0 \leq t \leq 1$ and $X_{i\ell} = 0$ on $M \backslash \pi_\ell(U_i)$.

Let $X = \sum_{i,\ell} X_{i\ell}$. The time-1 map $\varphi = \mathrm{e}^X$ has the required properties. ∎

Proof of lemma 2. To prove this lemma, for any given stochastic dynamical system with invariant probability p and dependence ratio f, and for any $\varepsilon > 0$, we have to construct a system with the same invariant probability and with dependence ratio $g \in N_\varepsilon(f)$ such that g is regular. The main technical difficulty, compared to the proof of the analogous statements for functions, is that we have 'fewer' perturbations to play with, since not every real-valued function on M^2 is a dependence ratio.

Changing transition probabilities. We begin by devising a way how we can change the transition probability of a process without changing its invariant probabilities. Recall equation (3):

$$p^{t+1}(x_{t+1}) = \int_M \tau(x_{t+1}|x_t) p^t(x_t) \, \mathrm{d}x_t.$$

Let $q(x_1, x_2)$ be a function such that

$$\int_M q(x_1, x_2) \, \mathrm{d}x_1 = \int_M q(x_1, x_2) \, \mathrm{d}x_2 = 0,$$

and such that $\tilde{\tau}(x_2|x_1) > 0$ for all $x_1, x_2 \in M$, where

$$\tilde{\tau}(x_2|x_1) = \tau(x_2|x_1) + \frac{q(x_1, x_2)}{p(x_1)}. \tag{10}$$

Then $\tilde{\tau}$ is also a transition probability density. Let p denote the unique invariant density satisfying $p(x_2) = \int \tau(x_2|x_1) p(x_1) \, \mathrm{d}x_1$ and $\int p(x) \, \mathrm{d}x = 1$. Note that then p satisfies also the integral equation

$$\int_M \tilde{\tau}(x_2|x_1) p(x_1) \, \mathrm{d}x_1 = \int_M \tau(x_2|x_1) p(x_1) \, \mathrm{d}x_1 + \int_M \frac{q(x_1, x_2)}{p(x_1)} p(x_1) \, \mathrm{d}x_1$$
$$= p(x_2).$$

This implies that p is also an invariant density of the stationary stochastic process with transition probability density $\tilde{\tau}$.

Construction of the perturbations. For given $\varepsilon > 0$, we shall construct a function $q : M^2 \to \mathbb{R}$ such that the perturbed dependence ratio

$$\tilde{f}(x_1, x_2) = \frac{\tilde{\tau}(x_2 | x_1)}{p(x_2)},$$

where $\tilde{\tau}$ is given by (10) is in $N_\varepsilon(\tau)$ and such that it is regular. This will be achieved by a transversality argument, for which we need a finite family of functions q_{ij}^k / p such that their differentials span the tangent space $\mathbb{R}^{2m}$ at every point $x \in M^2$. In the 'ordinary' case of the Morse classification of real valued functions on M, this is achieved by looking, for every $\xi \in M$, at functions of the form $\ell_{\xi,i}(x) = (x^i - \xi^i)\varphi\big(\|x - \xi\| / \delta\big)$, where $\varphi : \mathbb{R} \to \mathbb{R}$ is a smooth function such that $\varphi(t) = 1$ if $|t| < 1$ and $\varphi(t) = 0$ if $|t| > 2$. It is easily seen that this infinite family has the property that their differentials span $\mathbb{R}^m$ everywhere. By a compactness argument, we can find points $\xi_1, \cdots, \xi_n$, such this is still true for the functions $\ell_{\xi_k, i}$. Then one can show that a suitable perturbation $f + \sum a_i^k \ell_{\xi_k, i}$ of a given function f is regular. In our case, we have to modify this argument, as we need to preserve the property that the perturbed dependence ratio is still a dependence ratio.

Therefore, let $\varphi, \psi : \mathbb{R} \to \mathbb{R}$ be smooth functions such that

$$\varphi(t) = \begin{cases} 1 & \text{if} \quad |t| < 1, \\ 0 & \text{if} \quad |t| > 2, \end{cases}$$

and

$$\psi(t) = 0 \quad \text{if} \quad |t| < 1 \text{ or } |t| > 2, \qquad \psi(t) > 0 \quad \text{otherwise.}$$

Define functions $\ell_j^\xi : \mathbb{R}^m \to \mathbb{R}$, for $j = 0, 1, \cdots, m$ by

$$\ell_j^\xi(x) = \begin{cases} \varphi(\|x - \xi\| / \delta) + \beta_0 \psi(\|x - \xi\| / \delta) & \text{if} \quad j = 0, \\ (x^j - \xi^j)\varphi(\|x - \xi\| / \delta) + \beta_j \psi(\|x - \xi\| / \delta) & \text{if} \quad j \in \{1, \cdots, m\}; \end{cases}$$

here $\delta = \delta(\xi) > 0$ is sufficiently small such that the intersection of $B_{2\delta}(\xi) \backslash B_\delta(\xi)$ with M has positive measure. The constants $\beta_j = \beta_j(\xi)$ are determined by the requirement that $\int_M \ell_j^\xi \, \mathrm{d}x = 0$ for all j.

Let a positive function $P : M^2 \to \mathbb{R}$ be given, and for $\xi = (\xi_1, \xi_2) \in M^2$, define the functions (note that $x_1, x_2 \in M$):

$$L_{ij}^\xi(x_1, x_2) = \begin{cases} \dfrac{\ell_j^{\xi_1}(x_1)\ell_0^{\xi_2}(x_2)}{P(x_1, x_2)}, & i = 1, \\[3ex] \dfrac{\ell_j^{\xi_2}(x_2)\ell_0^{\xi_1}(x_1)}{P(x_1, x_2)}, & i = 2. \end{cases}$$

Denote by x_i^j the j'th coordinate of $x_i \in M$. We find that

$$\frac{\partial L_{ij}^\xi}{\partial x_{i'}^{j'}}(\xi_1, \xi_2) = \begin{cases} \dfrac{1}{P(\xi_1, \xi_2)} & \text{if} \quad i' = i,\ j' = j, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that the derivatives $DL_{ij}^\xi(\xi_1, \xi_2)$ are linearly independent. By continuity, there is a (relative) neighbourhood U_ξ of every point $\xi = (\xi_1, \xi_2)$ in M^2, such that the $2m$ vectors $DL_{ij}^\xi(x_1, x_2)$, $i = 1, 2$, $j = 1, \cdots, m$, are linearly independent for all $(x_1, x_2) \in U_\xi$.

Since M is compact, it is covered by a finite number of the set U_ξ, say U_{ξ_1}, $\cdots, U_{\xi_K}$. Set

$$q_{ij}^k(x_1, x_2) = P(x_1, x_2) L_{ij}^{\xi_k}(x_1, x_2).$$

Then q_{ij}^k / P is a finite collection of real-valued functions on M^2 such that their derivatives span $\mathbb{R}^{2m}$ at every point $(x_1, x_2) \in M^2$. Moreover

$$\int_M q_{ij}^k(x_1, x_2)\ \mathrm{d}x_1 = \int_M q_{ij}^k(x_1, x_2)\ \mathrm{d}x_2 = 0.$$

Let the stationary stochastic process defined by the transition probability $\tau(x_2|x_1)$ have invariant probability density p, invariant joint probability density $p^{1,2}(x_1, x_2)$ and dependence ratio

$$f(x_1, x_2) = \frac{p^{1,2}(x_1, x_2)}{p(x_1)p(x_2)} = \frac{\tau(x_2|x_1)}{p(x_2)}.$$

Let A be a sufficiently small open neighbourhood of 0 in $\mathbb{R}^{2mK}$ such that for all $a = (a_{ij}^k) \in A$, the function

$$\tilde{\tau}(x_2|x_1) = \tau(x_2|x_1) + \sum_{ijk} a_{ij}^k \frac{q_{ij}^k(x_1, x_2)}{p(x_1)}$$

is an everywhere positive transition probability density. Such a set A exists, since the transition probability density τ is assumed to be strictly positive everywhere on the compact manifold M^2; hence, if $\|a\|$ is small, then $\tilde{\tau}$ is strictly positive on M^2 as well. The dependence ratio of the new process is given by

$$g(a, x) = f(x) + \sum_{ijk} a_{ij}^k \frac{q_{ij}^k(x_1, x_2)}{p(x_1)p(x_2)}.$$

with $a = (a_{ij}^k) \in A$. If we set $P(x_1, x_2) = p(x_1)p(x_2)$, we have that $g(a, x)$ is a dependence ratio and that the derivatives of $q_{ij}^k(x_1, x_2)/p(x_1)p(x_2)$ span $\mathbb{R}^{2m}$ at every point of M^2.

27

The transversality argument. Recall the definition of transversality (see e.g. [5], definition 10.3.1, p. 83): if X and Y are smooth manifolds, W a smooth submanifold of Y, the map $f : X \to Y$ smooth, and $x \in X$, then f intersects W transversally at x, if either $f(x) \notin W$ or $f(x) \in W$ and $T_{f(x)}Y = T_{f(x)}W + \mathrm{d}f(x)\big(T_x X\big)$. More generally, we say that f intersects W transversally at $A \subset X$, if f intersects W transversally at x for every $x \in A$.

We have the theorem that if A, X and Y are smooth manifolds, W a smooth submanifold of Y and $f : A \times X \to Y$ a smooth map which intersects W transversally, then the set of points $a \in A$ for which $f_a = f(a,.) : X \to Y$ intersects W transversally is everywhere dense in A (see [5], theorem 10.3.3, p. 85).

The derivative $\mathrm{d}f$ of a function $f : M \to \mathbb{R}$ on an m-dimensional manifold M defines a section s of the cotangent bundle T^*M; in a sufficiently small neighbourhood U of a point in M, the restriction $T^*_U M$ of the bundle to U is isomorphic to $U \times \mathbb{R}^m$, and the section takes the form $s(x) = (x, Df(x))$. The zero section M_0 of T^*M, which is isomorphic to M, is locally of the form $U \times \{0\}$.

The section s is transversal to M_0 at a point $x \in M_0$, if either $s(x) \notin M_0$, or if

$$T_{(x,0)}T^*M = \mathrm{d}s(x)T_x M + T_{(x,0)}M_0 = (I, Hf(x))\mathbb{R}^m + \mathbb{R}^m \times \{0\}.$$

Note that this is equivalent to saying that s is transversal to M_0 everywhere if and only if the function f has only nondegenerate critical points. Such a function is called a *regular function* or a *Morse function*.

Consider now the function $g : A \times M \to \mathbb{R}$ constructed above and its associated map $s : A \times M \to T^*M$ given by $s(a, x) = (x, \mathrm{d}_x g(a, x))$. Note that s is transversal to M_0, since in local coordinates

$$\mathrm{d}s(a, x)T_{(a,x)}A \times M + T_{(x,0)}M_0$$
$$= \begin{pmatrix} 0 & I \\ D\frac{q_{ij}^k}{p(x_1)p(x_2)} & H_x g(a, x) \end{pmatrix} \mathbb{R}^{2mK} \times \mathbb{R}^m + \mathbb{R}^m \times \{0\},$$

and since by construction the derivatives $\mathrm{d}(q_{ij}^k(x_1, x_2)/p(x_1)p(x_2))$ span $\mathbb{R}^{2m}$ everywhere on M^2. By the theorem mentioned above, the set of $a \in A$, for which $g_a = g(a,.)$ is a Morse function, is an everywhere dense set in A.

Therefore, for every $\varepsilon > 0$, we can choose a so small that the dependence ratio $g = g_a$ is a Morse function and $g \in N_{\varepsilon/2}(f)$. It remains to show that by a second small perturbation, we can achieve that the perturbed dependence ratio is a regular dependence ratio.

The second perturbation. Note that since $g : M^2 \to \mathbb{R}$ is a Morse function, its critical points are isolated. Denote them by $\xi_1, \cdots, \xi_N \in M^2$.

We proceed inductively; for $k = 1$, we set $g_1 = g$. For $k > 1$, we assume that $g_k : M^2 \to \mathbb{R}$ is a dependence ratio that is a Morse function, $g_k \in N_{(1-2^k)\varepsilon}(f)$,

and that g_k has critical points $\tilde{\xi}_1, \cdots, \tilde{\xi}_{k-1}, \xi_k, \cdots, \xi_N$, such that the values $g_k(\tilde{\xi}_1)$ up to $g_k(\tilde{\xi}_{k-1})$ are different, and such that $\pi_\ell(\tilde{\xi}_i) \neq \pi_\ell(\tilde{\xi}_j)$ for $\ell = 1, 2$, if $1 \leq i < j \leq k - 1$ (recall that $\pi_\ell(x_1, x_2) = x_\ell$ is the projection on the ℓ'th coordinate).

We choose a neighbourhood $U \subset M^2$ of ξ_k such that ξ_k is the only critical point of g_k in U. Let $b \in \mathbb{R}^{2m}$ be such that $\langle b, Hg(\xi_k)^{-1} b \rangle \neq 0$, where $\langle x, y \rangle$ denotes the inner product of the vectors x and y; the inverse of $Hg_k(\xi_k)$ exists since g_k is nondegenerate in ξ_k; note that the set of vectors b that do not satisfy this condition is the union of the origin of $\mathbb{R}^{2m}$ and a smooth codimension-1 manifold.

Consider the function

$$h_t(x_1, x_2) = h(t, x_1, x_2) = g_k(x_1, x_2) - t \sum_{ij} b_{ij} L_{ij}^{\xi_k}(x_1, x_2).$$

The critical points of h_t are determined by the equation

$$0 = Dh_t(x_1, x_2) = Dg_k(x_1, x_2) - t \sum_{ij} b_{ij} DL_{ij}^{\xi_k}(x_1, x_2).$$

We solve $x = (x_1, x_2)$ as a function of t in a neighbourhood of $t = 0$ and $(x_1, x_2) = \xi_k$, using the fact that $Hg_k(\xi_k)$ is invertible. For the solution $x(t) = (x_1(t), x_2(t))$, we find that

$$\frac{\mathrm{d}x}{\mathrm{d}t}(0) = \frac{1}{p(\pi_1 \xi_k) p(\pi_2 \xi_k)} Hg_k(\xi_k)^{-1} b. \tag{11}$$

Note that by the assumption on b, this derivative is nonzero. We restrict the possible choice of b further by requiring that

$$\frac{\mathrm{d}x_\ell}{\mathrm{d}t}(0) = D\pi_\ell \frac{\mathrm{d}x}{\mathrm{d}t}(0) = \frac{1}{p(\pi_1 \xi_k) p(\pi_2 \xi_k)} D\pi_\ell Hg_k(\xi_k)^{-1} b \neq 0.$$

Moreover, setting $v(t) = h_t(x(t))$, we have

$$\frac{\mathrm{d}v}{\mathrm{d}t}(t) = -\sum_{ij} b_{ij} L_{ij}^{\xi_k} + D_x h_t(x) = -\sum_{ij} b_{ij} L_{ij}^{\xi_k},$$

and

$$\frac{\mathrm{d}^2 v}{\mathrm{d}t^2}(\xi_k) = -\frac{1}{p(\pi_1 \xi_k) p(\pi_2 \xi_k)} \langle b, Hg_k(\xi_k)^{-1} b \rangle \neq 0. \tag{12}$$

There are only finitely many values of t for which $v(t)$ is equal to the critical values $g_k(\xi_i)$, or for which the projection $\pi_\ell(x(t))$ coincides with $\pi_\ell(\xi_i)$ for some $1 \leq i < k$ and some $\ell = 1, 2$. From equations (11) and (12) it follows that the set of values of t avoiding these special values is everywhere dense in a neighbourhood of $t = 0$.

Choose t_k in this set such that the function

$$g_{k+1} = g_k - t_k \sum_{ij} b_{ij} L_{ij}^{\xi_k}$$

satisfies $g_{k+1} \in N_{(1-2^{k+1})\varepsilon}(f)$, and put $\tilde{\xi}_k = x(t_k)$. Then the function g_{k+1} satisfies the induction hypothesis. We conclude that the function g_{N+1} is a regular dependence ratio, and $g_{N+1} \in N_\varepsilon(f)$. This finishes the proof of the lemma. ∎

Acknowledgements

References

[1] I.S. Abramson, *On bandwidth variation in kernel estimates - a square root law*, Annals of Statistics **10** (1982), 1217–1223.

[2] L. Arnold, *Random dynamical systems*, Springer, Heidelberg, 1998.

[3] L. Cobb, *Stochastic catastrophe models and multimodal distributions*, Behavioral Science **23** (1978), 360–374.

[4] M.A.H. Dempster, I.V. Evstigneev, and K.R. Schenk-Hoppé, *Exponential growth of fixed-mix strategies in stationary asset markets*, Finance and Stochastics **7** (2003), 263–276.

[5] B.A. Dubrovin, A.T. Fomenko, and S.P. Novikov, *Modern Geometry — Methods and Applications. Part II: The Geometry and Topology of Manifolds*, Graduate Texts in Mathematics, vol. 104, Springer, New York, 1985.

[6] M.W. Hirsch, *Differential Topology*, Graduate Texts in Mathematics, vol. 33, Springer, New York, 1976.

[7] P. W. Holland and Y. J. Wang, *Dependence function for bivariate densities*, Communications in Statistics **A 16** (1987), 863–876.

[8] M. C. Jones, *The local dependence function*, Biometrika **83** (1996), 899–904.

[9] S. Nadarajah, K. Mitov, and S. Kotz, *Local dependence functions for extreme value distributions*, Journal of Applied Statistics **30** (2003), 1081–1100.

[10] John C. Oxtoby, *Measure and category. A survey of the analogies between topological and measure spaces. 2nd ed.*, Graduate Texts in Mathematics, vol. 2, Springer, 1980.

[11] A. Ploeger, H. L. J. van der Maas, and P. Hartelman, *Catastrophe Analysis of Switches in the Perception of Apparent Motion*, Psychonomic Bulletin & Review **9** (2002), 26–42.

[12] G.R. Terrell and D.W. Scott, *Variable kernel density estimation*, Annals of Statistics **20** (1992), 1236–1265.

[13] René Thom, *Structural stability and morphogenesis. An outline of a general theory of models*, W. A. Benjamin, Reading, Massachusetts, 1975.

[14] E.-J. Wagenmakers, P.C.M. Molenaar, R.P.P.P. Grasman, P.A.I. Hartelman, and H.L.J. van der Maas, *Transformation invariant stochastic catastrophe theory*, Physica D **211** (2005), 263–276.

[15] E.C. Zeeman, *Stability of dynamical systems*, Nonlinearity **1** (1988), 115–155.