

14 Dimension of Attractors Estimation

How can we prove that all the attractors of an infinite-dimensional system belong to a finite-dimensional manifold? How can we estimate the dimensions of attractor? There exist two methods to obtain such estimations.

First, if we find that *k-dimensional volumes are contracted* due to dynamics, then (after some additional technical steps) we can claim that the *Hausdorff dimension* of the maximal attractor is less than *k*.

Second, if we find a representation of our system as a nonlinear kinetic system with *conservation of supports* of distributions, then (again, after some additional technical steps) we can state that the asymptotics is finite-dimensional. This conservation of support has a *quasi-biological interpretation, inheritance* (if a gene was not presented initially in an isolated population without mutations, then it cannot appear at later time). The finite-dimensional asymptotic demonstrates effects of “*natural*” selection.

In this chapter we describe these approaches.

14.1 Lyapunov Norms, Finite-Dimensional Asymptotics and Volume Contraction

In a general case, it is impossible to prove the existence of a global Lyapunov function on the basis of local data. We can only verify or falsify the hypothesis about a given function, is it a global Lyapunov function, or is it not. On the other hand, there exists a more strict stability property which can be verified or falsified (in principle) with local data analysis. This is a Lyapunov norm existence.

A norm $\|\bullet\|$ is the *Lyapunov norm* for the system (13.1), if for any two solutions $x^{(1)}(t)$, $x^{(2)}(t)$, $t \geq 0$, the function $\|x^{(1)}(t) - x^{(2)}(t)\|$ is non-increasing in time.

A linear operator A is *dissipative* with respect to a norm $\|\bullet\|$, if $\exp(At)$ ($t \geq 0$) is a semigroup of contractions: $\|\exp(At)x\| \leq \|x\|$ for any x and $t \geq 0$. A family of linear operators $\{A_\alpha\}_{\alpha \in K}$ is *simultaneously dissipative*, if all operators A_α are dissipative with respect to some norm $\|\bullet\|$ (it should be stressed that in this definition one requires *one* norm for *all* A_α , $\alpha \in K$). The mathematical theory of simultaneously dissipative operators for finite-dimensional spaces was developed in [299–303].

Let the system (13.1)

$$\frac{dx}{dt} = J(x) + J_{ex}(x, t), x \in U$$

be defined in a convex set $U \subset E$, and A_x be Jacobi operator at the point $x: A_x = D_x(J(x) + J_{ex}(x))$. This system has a Lyapunov norm, if the family of operators $\{A_x\}_{x \in U}$ is simultaneously dissipative. If one can choose such $\varepsilon > 0$ that for all $A_x, t > 0$, any vector z , and this Lyapunov norm

$$\| \exp(A_x t) z \| \leq \exp(-\varepsilon t) \| z \| ,$$

then for any two solutions $x^{(1)}(t), x^{(2)}(t), t \geq 0$ of equations (13.1)

$$\| x^{(1)}(t) - x^{(2)}(t) \| \leq \exp(-\varepsilon t) \| x^{(1)}(0) - x^{(2)}(0) \| .$$

The existence of the Lyapunov norm is a very strong restriction on non-linear systems, and such systems are not widespread in applications. But if we go from the distance contraction to the contraction of the k -dimensional volumes ($k = 2, 3, \dots$) [311], the situation changes dramatically. There exist many kinetic systems with a monotonous contraction of the k -dimensional volumes for sufficiently large k (see, for example, [173, 317, 318, 334]). Let $x(t), t \geq 0$ be a solution of equation (13.1). Let us write the first approximation equation for small variation of $x(t)$:

$$\frac{d\Delta x}{dt} = A_{x(t)} \Delta x . \tag{14.1}$$

This is a linear system with coefficients dependent on t . Let us study how the system (14.1) changes the k -dimensional volumes. For k -dimensional parallelepiped with the edges $x^{(1)}, x^{(2)}, \dots, x^{(k)}$ we can define an element of the k th exterior power, the oriented volume:

$$x^{(1)} \wedge x^{(2)} \wedge \dots \wedge x^{(k)} \in \underbrace{E \wedge E \wedge \dots \wedge E}_k = \wedge^k E ,$$

(this is an antisymmetric tensor or k -vector). A norm in the k th exterior power of the space E is a measure of the k -dimensional volumes (one of the possible measures). Dynamics of volumes induced by the system (14.1) is given by equations

$$\begin{aligned} \frac{d}{dt} (\Delta x^{(1)} \wedge \Delta x^{(2)} \wedge \dots \wedge \Delta x^{(k)}) &= (A_{x(t)} \Delta x^{(1)}) \wedge \Delta x^{(2)} \wedge \dots \wedge \Delta x^{(k)} \\ &+ \Delta x^{(1)} \wedge (A_{x(t)} \Delta x^{(2)}) \wedge \dots \wedge \Delta x^{(k)} + \dots \\ &+ \Delta x^{(1)} \wedge \Delta x^{(2)} \wedge \dots \wedge (A_{x(t)} \Delta x^{(k)}) \\ &= (\wedge_D^k A_{x(t)}) (\Delta x^{(1)} \wedge \Delta x^{(2)} \wedge \dots \wedge \Delta x^{(k)}) . \end{aligned} \tag{14.2}$$

Here $\wedge_D^k A_{x(t)}$ are operators of the *induced differential action* of $A_{x(t)}$ on the k th exterior power of E . Again, a decrease of $\|\Delta x^{(1)} \wedge \Delta x^{(2)} \wedge \dots \wedge \Delta x^{(k)}\|$ in time is equivalent to dissipativity of all the operators $\wedge_D^k A_{x(t)}$, $t \geq 0$ in the norm $\|\bullet\|$. Existence of such a norm for all $\wedge_D^k A_{x(t)}$ ($x \in U$) is equivalent to the contraction of the volumes of all parallelepipeds due to first approximation system (14.1) for any solution $x(t)$ of equations (13.1). We shall call such a system the *k-contraction*. If one can choose such $\varepsilon > 0$ that for all A_x ($x \in U$), any vector $z \in E \wedge E \wedge \dots \wedge E$, and this norm,

$$\|\exp(\wedge_D^k A_{x(t)} t) z\| \leq \exp(-\varepsilon t) \|z\|,$$

then the volumes of the parallelepipeds contract exponentially as $\exp(-\varepsilon t)$.

For such systems we can estimate the Hausdorff dimension of the attractor (under some additional technical conditions about solutions boundedness): it cannot exceed k . It is necessary to stress here that this estimation of the Hausdorff dimension does not solve the problem of construction of the invariant manifold containing this attractor, and one needs special technique and additional restriction on the system in order to obtain this manifold (see [318, 342, 346, 347]).

Let us remind here the definition of the *Hausdorff dimension* of metric space (subset of a normed space, for example). Let X be a metric space, and d be a number. The *d-dimensional Hausdorff measure* of X , $H^d(X)$, is the infimum of positive numbers y such that for every $r > 0$, X can be covered by a countable family of closed balls, each of diameter less than r , such that the sum of the d th powers of their diameters is less than y . Note that $H^d(X)$ may be infinite, and d need not be an integer. The Hausdorff dimension $D_H(X)$ of X is the infimum of such $d \geq 0$ that the $H^d(X) = 0$:

$$D_H(X) = \inf\{d \geq 0 | H^d(X) = 0\}.$$

The simplest way for construction of the slow invariant manifold becomes available for systems with a dominance of the linear part in higher dimensions. Let an infinite-dimensional system have a form: $\dot{u} + Au = R(u)$, where the operator A is self-adjoint, and has discrete spectrum λ_i ($\lambda_1 < \lambda_2 < \dots$, $\lambda_i \rightarrow \infty$) with sufficiently big gaps between the eigenvalues λ_i , and $R(u)$ is continuous. Let E_i be the eigenspace of A for the eigenvalue λ_i , and all the E_i are finite-dimensional spaces. One can build the slow manifold as the graph over the space $\oplus_{i=1}^k E_i$ for some k . Indeed, let the basis consists of eigenvectors of A . In this basis $\dot{u}_i = -\lambda_i u_i + R_i(u)$, and it seems plausible that, for some k and sufficiently big i , the functions $u_i(t)$ exponentially fast tend to $u_i(u_1(t), \dots, u_k(t))$, if $R_i(u)$ are bounded and continuous in a suitable sense. Here $u_i(u_1, \dots, u_k)$ are some smooth functions that describe the manifold with internal coordinates u_1, \dots, u_k .

Variants of rigorous theorems about systems with such a dominance of the linear part in higher dimensions may be found in literature (see, for example, the textbook [100]). Even if all the sufficient conditions hold, efficient

computation of these manifold remains the problem, and various strategies of calculations are proposed: from the Euler method for the manifold correction [25] to various algorithms of discretization [13, 105, 343].

The simplest condition of simultaneous dissipativity for the family of operators $\{A_x\}$ can be obtained in a following way: let us take a norm $\|\bullet\|$. If all operators A_x are dissipative with respect to this norm, then the family A_x is (obviously) simultaneously dissipative in this norm. So, we can verify or falsify a hypothesis about the simultaneous dissipativity for a given norm. Simplest examples are provided by the quadratic and l^1 norms.

For the quadratic norm associated with a scalar product $\langle \cdot | \cdot \rangle$ the dissipativity of the operator A is equivalent to nonpositivity of the spectrum of the operator $A + A^+$, where A^+ is the adjoint to A operator with respect to scalar product $\langle \cdot | \cdot \rangle$.

For the l^1 norm with weights $\|x\| = \sum_i w_i |x_i|$, $w_i > 0$, the condition of the operator's A dissipativity in this norm is the weighted diagonal dominance for columns of the matrix $A = (a_{ij})$:

$$a_{ii} < 0, w_i |a_{ii}| \geq \sum_{j, j \neq i} w_j |a_{ji}|.$$

For the exponential contraction, it is necessary and sufficient that some gap exists in the dissipativity inequalities:

- For the quadratic norm,

$$\sigma(A + A^+) < \varepsilon < 0,$$

where $\sigma(A + A^+)$ is the spectrum of $A + A^+$;

- For the l^1 norm with weights,

$$a_{ii} < 0, w_i |a_{ii}| \geq \sum_{j, j \neq i} w_j |a_{ji}| + \varepsilon, \varepsilon > 0.$$

Sufficient conditions of simultaneous dissipativity can have a different form (not only the form of dissipativity checking with respect to a given norm) [300–303], but the problem of necessary and sufficient conditions in a general case remains open.

The dissipativity conditions for operators $\wedge_D^k A_x$ of the differential induced action of A_x on the k th exterior power of E have a similar form. If we know the spectrum of $A_x + A_x^+$, then it is easy to find the spectrum of $\wedge_D^k A_x + (\wedge_D^k A_x)^+$. For example, for simple discrete spectrum each eigenvalue of this operator is a sum of k distinct eigenvalues of $A_x + A_x^+$.

A basis of the k th exterior power of E can be constructed from the basis $\{e_i\}$ of E : it is

$$\{e_{i_1 i_2 \dots i_k}\} = \{e_{i_1} \wedge e_{i_2} \wedge \dots \wedge e_{i_k}\}, i_1 < i_2 < \dots < i_k.$$

For the l^1 norm with weights in the k th exterior power of E the set of weights is $\{w_{i_1 i_2 \dots i_k} > 0, i_1 < i_2 < \dots < i_k\}$. The norm of a vector z is

$$\|z\| = \sum_{i_1 < i_2 < \dots < i_k} w_{i_1 i_2 \dots i_k} |z_{i_1 i_2 \dots i_k}|.$$

The dissipativity conditions for operators $\wedge_D^k A_{x(t)}$ of the induced differential action of A in the l^1 norm with weights have the form:

$$\begin{aligned} a_{i_1 i_1} + a_{i_2 i_2} + \dots + a_{i_k i_k} &< 0, \\ w_{i_1 i_2 \dots i_k} |a_{i_1 i_1} + a_{i_2 i_2} + \dots + a_{i_k i_k}| &\geq \sum_{l=1}^k \sum_{j, j \neq i_1, i_2, \dots, i_k} w_{i_1 i_2 \dots i_k}^{l,j} |a_{j i_l}| \\ \text{for any } i_1 < i_2 < \dots < i_k, & \end{aligned} \quad (14.3)$$

where $w_{i_1 i_2 \dots i_k}^{l,j} = w_I$, the multiindex I consists of indexes i_p ($p \neq l$), and j .

For infinite-dimensional systems the problem of volume contraction and Lyapunov norms for exterior powers of E consists of three parts: the geometrical part that concerns the choice of norm for simultaneous dissipativity of operator families, the topological part that concerns the topological non-equivalence of the constructed norms, and estimation of the bounded set (the so-called *absorbing set*) containing a compact attractor. This appropriate a priori estimations of the bounded convex positively invariant set $V \subset U$ where the compact attractor is situated may become a difficult problem.

The estimation of the attractor dimension based on Lyapunov norms in the exterior powers is rather rough. This is a local estimation. Refined estimations are based on global Lyapunov exponents (Lyapunov or Kaplan-Yorke dimension [308, 309]). There are many different measures of the dimension [307, 310], and much effort is invested towards better estimates of various dimensions [345].

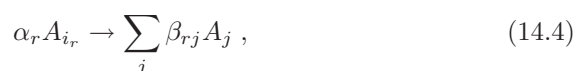
Estimations of the dimension of attractors was given for various systems: from the Navier-Stokes hydrodynamics [316] to the climate dynamics [312]. An introduction and a review of many results is presented in the book [318]. The local estimations remain the main tool for the estimation of the attractors dimension, because global estimations for complex systems are much more complicated and often just unattainable because of the computational complexity.

14.2 Examples: Lyapunov Norms for Reaction Kinetics

In this section we consider the reaction kinetics systems which obey the mass action law. The direct and reverse reactions will be represented in the stoichiometric equations separately, and the reaction mechanisms under consideration can include an elementary reaction without its reverse reaction.

Neither detailed balance, nor other dissipation requirements are presumed. Such schemes appear, for example, in modeling of catalytic reactions, where only part of the components are included into equations, and concentrations of other components are used as parameters [81].

The simplest class of nonlinear kinetic (open) systems with Lyapunov norms was described in the paper [304]. These are reaction systems without interactions of various substances. The stoichiometric equation of each elementary reaction has a form



where r enumerates reactions, i_r is defined for each reaction α_r , β_{rj} are non-negative stoichiometric coefficients (usually they are integer), A_i are symbols of substances.

In the left hand part of equation (14.4) there is one initial reagent, though $\alpha_r > 1$ is possible (there may be several copies of A_{i_r} , for example $3A \rightarrow 2B + C$). This explains the notion of the reaction without interaction of different substances.

For the mass action law (2.67) kinetic equations for the kinetic scheme (14.4) have a form

$$\frac{dc_i}{dt} = - \sum_{r: i=i_r} \alpha_r k_r c_i^{\alpha_r} + \sum_r \beta_{ri} k_r c_{i_r}^{\alpha_r}, \quad k_r > 0. \quad (14.5)$$

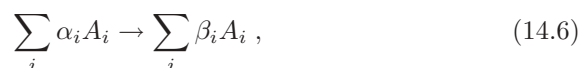
Let there exist a positive balance for the reaction scheme (14.4):

$$b_{i_r} \alpha_r = \sum_j b_j \beta_{rj}$$

for some $b_i > 0$ and all r . In this case kinetic equations for the reaction system (14.4) have a Lyapunov norm [304]. This is the weighted l^1 norm: $\|x\| = \sum_i b_i |x_i|$, where $b_i > 0$ are the coefficients from the linear conservation law $\sum_i b_i c_i = \text{const}$. There exists no quadratic Lyapunov norm for general reaction systems without interaction of different substances.

Let us call a reaction mechanism *dissipative*, if it allows a universal Lyapunov norm¹, that is, the Lyapunov norm which is independent of reaction rate constants.

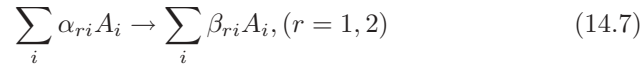
The reaction mechanism with only one elementary reaction,



is dissipative if and only if for any i if $\alpha_i > 0$, then $\alpha_i > \beta_i$.

¹ For the motion on the planes with given values of the linear conservation laws.

The reaction mechanism with two elementary reaction without autocatalysis ($\alpha_{ri} \times \beta_{ri} \equiv 0$),



is *not* dissipative if and only if for two of components A_i, A_j at least one of the following conditions holds:

1. Branching

$$\alpha_{1i}, \beta_{1j}, \alpha_{2j}, \beta_{2i} > 0, \frac{\beta_{1j}}{\alpha_{1i}} \frac{\beta_{2i}}{\alpha_{2j}} > 1, \quad (14.8)$$

2. Loss disproportion

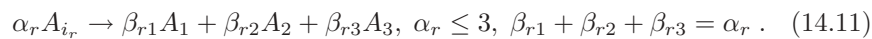
$$\alpha_{1i}, \alpha_{1j}, \alpha_{2i}, \alpha_{2j} > 0, \frac{\alpha_{1i}}{\alpha_{1j}} \neq \frac{\alpha_{2i}}{\alpha_{2j}}, \quad (14.9)$$

3. Gain–loss connection

$$\alpha_{1i}, \alpha_{1j}, \alpha_{2i}, \beta_{2j} > 0, \frac{\alpha_{1i}}{\alpha_{1j}} \frac{\beta_{2j}}{\alpha_{2i}} > 1. \quad (14.10)$$

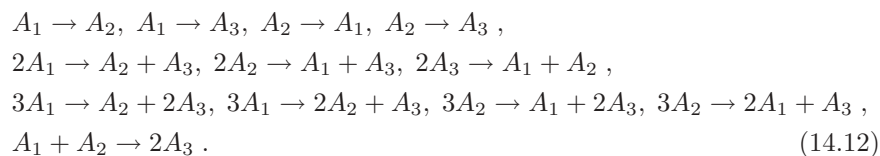
These elementary *obstacles to dissipativity* allow to prove the non-dissipativity of many reaction mechanisms.

It is possible to describe all the dissipative reaction mechanisms. For example, for three components A_1, A_2, A_3 with the linear conservation law $c_1 + c_2 + c_3 = \text{const}$, for bounded stoichiometric numbers ($\alpha_{ri}, \beta_{ri} \leq 3$), and without autocatalysis ($\alpha_{ri} \times \beta_{ri} \equiv 0$) there are three such maximal mechanisms, and any other dissipative mechanism is a subset of one of these maximal schemes (may be after permutation of the components) [305]. The first maximal scheme consists of all the reactions without interaction of different substances and with given conservation law:



The corresponding Lyapunov norm is $\|c\| = |c_1| + |c_2| + |c_3|$.

The second maximal scheme includes a reaction with interaction of different substances, $A_1 + A_2 \rightarrow 2A_3$:



The corresponding norm is $\|c\| = |c_1| + |c_2| + |c_1 + c_2 + c_3|$.

The third maximal scheme includes another reaction with interaction of different substances, $A_1 + 2A_2 \rightarrow 3A_3$:

$$\begin{aligned}
& A_1 \rightarrow A_2, A_1 \rightarrow A_3, A_2 \rightarrow A_3, \\
& 2A_1 \rightarrow A_2 + A_3, 2A_2 \rightarrow A_1 + A_3, \\
& 3A_1 \rightarrow A_2 + 2A_3, 3A_1 \rightarrow 2A_2 + A_3, 3A_2 \rightarrow A_1 + 2A_3, 3A_3 \rightarrow A_1 + 2A_2, \\
& A_1 + 2A_2 \rightarrow 3A_3.
\end{aligned} \tag{14.13}$$

The corresponding norm is $\|c\| = 2|c_1| + |c_2| + |c_1 + c_2 + c_3|$.

14.3 Examples: Infinite-Dimensional Systems With Finite-Dimensional Attractors

In this section we list some of known examples of infinite-dimensional systems that are k -contractions (that is, they contract k -dimensional volume for some k) and have finite-dimensional attractors.

The most celebrated example are the *Navier–Stokes equations* for incompressible fluid in a bounded domain Ω of R^3 or R^2 :

$$\begin{aligned}
& \operatorname{div} \mathbf{u} = 0, \\
& \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}, \nabla) \mathbf{u} - \nu \Delta \mathbf{u} + \nabla p = \mathbf{f},
\end{aligned} \tag{14.14}$$

where $\mathbf{u}(\mathbf{x}, t)$ is the velocity field, ν is the kinematic viscosity, $\Delta = \nabla^2$ is the Laplace operator, (\cdot, \cdot) is the standard inner product, p is the pressure, \mathbf{f} is the volume force (normalized to unit density).

The two main cases are studied, the flow with the Dirichlet boundary conditions $\mathbf{u} = 0$ on $\partial\Omega$, or the flow with periodic boundary conditions.

For three-dimensional flows it is possible to prove conditional theorems of such a kind [313]: If X is a bounded attractor, then

$$D_H(X) \leq \operatorname{const} \operatorname{Re}^3, \tag{14.15}$$

where const is an universal constant and Re is the Reynolds number. The usual definition of Re is

$$\operatorname{Re} = \frac{LU}{\nu}, \tag{14.16}$$

where L and U are the reference length and reference velocity.

There is no natural “typical velocity” in a turbulent flow and therefore there is no obvious unique choice of the Reynolds number. The different versions of the Reynolds numbers are discussed, for instance, in [313, 314].

The classical physical estimate of the number of degrees of freedom for flows with large Reynolds numbers is $\sim \operatorname{Re}^{9/4}$ [315]. This estimate was obtained mathematically rigorously for the homogeneous decaying turbulence [316]. That is, for the Navier–Stokes equations without a force \mathbf{f} and after a special time-dependent transformation of space and time scales.

For two-dimensional flows it is possible to prove the existence of attractors, because in this case the situation with existence and uniqueness of solutions is much more clear. There is a family of estimates of attractor dimension [318]. The simplest and at the same time most important form of such an estimate is

$$D_H(X) \leq \text{const} (G + 1), \quad (14.17)$$

where G is the generalized Grashof number:

$$G = \frac{L^2 \|\mathbf{f}\|}{\nu^2}, \quad (14.18)$$

$\|\mathbf{f}\|$ is the L^2 -norm of \mathbf{f} in Ω , L is the reference length. Of course, this definition allows some useful variations, as well, as the Reynolds number.

For special classes of solutions these estimates could be made more precise. For example, for the boundary and pressure-gradient driven incompressible fluid flows in elongated two-dimensional channels the following estimates of the dimension of the attractor for the solutions of the Navier–Stokes equations are obtained [320]. For boundary driven shear flows and flux driven channel flows the upper bounds for the degrees of freedom was found in the form $c\alpha \text{Re}^{3/2}$ where c is a universal constant, α denotes the aspect ratio of the channel (length/width), and Re is the Reynolds number based on the channel width and the imposed “outer” velocity scale. For fixed pressure gradient driven channel flows an upper bound of form $c' \text{Re}^2$ was obtained, where c' is another universal positive constant and the Reynolds number is based on a velocity defined by the infimum, over all possible trajectories, of the time averaged mass flux per unit channel width.

The estimates of the global attractor dimension were obtained for so-called *generalized Navier–Stokes* equations characterized by polynomial dependence between the stress tensor and the symmetric velocity gradient [319].

The maximal attractor X for the *complex Ginzburg–Landau equation* periodic on the interval $[0, 1]$

$$\frac{\partial A}{\partial t} = RA + (1 + i\nu) \frac{\partial^2 A}{\partial x^2} - (1 + i\mu)A|A|^2 \quad (14.19)$$

has a finite dimension [321, 322]:

– For $|\mu| \leq \sqrt{3}$

$$D_H(X) < 2\sqrt{3}(R^{1/2}/2\pi) + 1;$$

– For $|\mu| > \sqrt{3}$

$$D_H(X) < (\sqrt{3}/\pi)|\mu|R + 3|\mu|^{1/2}R^{1/2}/2\pi + 2.$$

The *Kuramoto–Sivashinsky equation* was invented in order to describe waves in chemically reacting media:

$$\frac{\partial c}{\partial t} = -\nu\Delta^2 c - \Delta c - \frac{1}{2}(\nabla c)^2. \quad (14.20)$$

where $\nu > 0$ and Δ is the Laplace operator. The term $-\Delta c$ in the right hand side of this equation describes the negative diffusion, and the solutions remain regular due to the term $-\nu\Delta^2 c$.

The global dynamical properties of the Kuramoto–Sivashinsky equation were studied in [323]. In particular, in the one-dimensional case for *even* solutions with the L -periodic boundary conditions, the following estimate for the dimension of the attractor X which includes all bounded attractors was obtained:

$$D_H(X) \leq 5.4\tilde{L}^{13/8}, \quad (14.21)$$

where $\tilde{L} = L/(2\pi\sqrt{\nu})$ is the dimensionless size of the box (or of the “pattern cell”).

For the two- and three-dimensional cases the estimations of attractor dimension include an a priori estimate of the L^2 norm of the c gradient $R = \overline{\lim}_{t \rightarrow \infty} \|\nabla c(t)\|$ [323].

Dynamics of the *nonlocal Kuramoto–Sivashinsky equation* (in the one-dimensional case) was studied in [324]

$$\frac{\partial c}{\partial t} = -\frac{\partial^4 c}{\partial x^4} - \frac{\partial^2 c}{\partial x^2} - c\frac{\partial c}{\partial x} - \alpha H\left(\frac{\partial^3 c}{\partial x^3}\right), \quad (14.22)$$

where $\alpha > 0$, and

$$H(f) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(\xi)}{x - \xi} d\xi, \quad (14.23)$$

(the integral is understood in the sense of the Cauchy principle value). This equation arises in the modeling of the flow of a thin film of viscous liquid down an inclined plane in electric field.

For large α and l the obtained attractor dimension estimate is

$$D_H(X) \leq O(\alpha^{3/2}\tilde{L}^{3/2} + \tilde{L}^2). \quad (14.24)$$

The *Cahn–Hilliard equation* describes the evolution of a conserved concentration field during phase separation:

$$\frac{\partial c}{\partial t} = \Delta(-\nu\Delta c + f(c)), \quad (14.25)$$

where $f(c)$ is a polynomial of an odd order. Usually, $f(c) = -\alpha c + \beta c^3$, $\alpha, \beta > 0$. Like the Kuramoto–Sivashinsky equation, the Cahn–Hilliard equation contains the regularizing term $-\nu\Delta^2 c$ and the negative diffusion term $-\alpha\Delta c$ in the right hand side. The operator $-\nu\Delta^2$ effectively suppresses the short waves (with the logarithmic decrement $\sim k^4$ for the wave vector k); it allows to prove the existence of the finite-dimensional attractors and to estimate their dimensions for various hypothesis about f and boundary conditions [325–329]. The Cahn–Hilliard equation can be presented as a gradient

system that minimizes the free energy (possibly, to a local minimum). Hence, the limiting behavior of this system is rather simple; it always tends to a stationary point (to the stable or metastable equilibrium).

It should be stressed that the maximal attractor (4.1)

$$X = \bigcap_{t \geq 0} T_t(Y),$$

where Y is the positively invariant absorbing set, and T_t is the map of the phase flow (the shift operator for the time t), estimates the limit behavior from above and often does not satisfy the intuitive expectation (that is, it is too large). For example, if the absorbing set Y includes a saddle point then the corresponding maximal attractor X includes the whole unstable manifold of this point, and the dimension of X is not less than the dimension of this unstable manifold. For instance, when all the solutions tend to fixed points from a finite set, the dimension of maximal attractor might be nonzero and rather large. Therefore, the study of the *structure* of the attractor becomes important. For the *viscous Cahn–Hilliard equation*

$$(1 - \alpha) \frac{\partial c}{\partial t} = \Delta \left(-\nu \Delta c + f(c) + \alpha \frac{\partial c}{\partial t} \right), \quad (14.26)$$

with usual no-flow boundary conditions the structure of the global attractor was studied in [330] (for the one-dimensional case). The dimension of the unstable manifolds was calculated for all stationary states. In the unstable case, the flow on the global attractor is shown to be semi-conjugate to the flow of the global attractor of the *Chaffee–Infante equation*

$$\frac{\partial c}{\partial t} = \frac{\partial^2 c}{\partial x^2} + \lambda^2(c - c^3) \quad (14.27)$$

with zero boundary conditions. The connection between these equations is in accordance with their physical sense [331, 332].

In the *reaction–diffusion equations* the only source of instabilities can be the reaction part. The diffusion suppresses short waves (with the logarithmic decrement $\sim k^2$ for wave vector k). Let us consider the following system in a bounded domain $\Omega \subset R^n$ with the zero or non-flow boundary conditions [333, 335]:

$$\frac{\partial c}{\partial t} = \nu d \Delta c - f(c) + g(x), \quad (14.28)$$

where $c(x, t) = (c^1, \dots, c^N)$, $\nu > 0$, $d = d_{ij}$ is a real $N \times N$ matrix with positive symmetric part $(d + d^+)/2 \geq 1$, $f(c) = (f^1(c), \dots, f^N(c))$, and $g(x) = (g^1(x), \dots, g^N(x))$. Here, $f(c)$ is the reaction part, and $g(x)$ describes the external sources and sinks.

Let $f(c)$ be a smooth map, $D_c f(c) = (\partial f^i / \partial c^j)$, and the following conditions hold

$$\sum_i \gamma_i |c^i|^{p_i} - C_1 \leq (c, f(c)) ; \tag{14.29}$$

$$\sum_i |f^i(c)|^{q_i} \leq C_2 \left(\sum_i |c^i|^{p_i} + 1 \right) ; \tag{14.30}$$

$$(D_c f(c)w, w) \geq -C_3 |w|^2 , \tag{14.31}$$

for some $C_1, C_2, C_3 > 0$, $\gamma_i > 0$, $p_i \geq 2$, $1/p_i + 1/q_i = 1$, any c and all $w \in R^N$. Then the dimension of the attractor X for (14.28) admits the following estimate:

$$D_H(X) \leq N|\Omega| \left(\frac{C_3}{\nu c_0(\Omega)} \right)^{n/2} , \tag{14.32}$$

where n is the space dimension, $|\Omega|$ is the measure of $\Omega \subset R^n$, and $c_0(\Omega)$ depends on the shape of Ω only, that is, $c_0(\lambda\Omega) = c_0(\Omega)$ for all $\lambda > 0$. The constant C_3 estimates the possible increment of the kinetic trajectories instability from above.

The reaction–diffusion systems remain the basic example of the finite-dimension volumes contractions, because of transparent structure: the diffusion generates contraction (with the logarithmic decrement $\sim k^2$ for wave vector k), and the reaction can cause finite-dimensional instabilities. It is well-studied, see, for instance, [318, 333–336].

For the *dissipative wave equations* the estimates of attractor dimensions were found in [333, 337–339] and discussed in [318, 335]. Let us consider the following equation in a bounded domain $\Omega \subset R^n$

$$\frac{\partial^2 u}{\partial t^2} + \gamma \frac{\partial u}{\partial t} = \Delta u - f(u) + g(x) \tag{14.33}$$

with zero boundary conditions. This equation contains the damping term $\gamma \frac{\partial u}{\partial t}$ with $\gamma > 0$. The essential difference between (14.33) and the reaction–diffusion equations (14.28) is in the character of dissipation. According to the dissipative wave equations short waves decay as $\exp(-\gamma t)$, the logarithmic decrement has the constant limit for $k^2 \rightarrow \infty$. It is the same situation, as for the Grad equations (see Chap. 8), or for the model kinetic equations (see Chap. 2).

Let $0 < \lambda_1 \leq \lambda_3 \leq \lambda_4 \leq \dots$ be the eigenvalues of the operator $-\Delta$ in Ω with zero boundary conditions (in the order of their values). Assume that $g \in L^2(\Omega)$ and the smooth function $f(u)$ satisfies the following conditions for any u

$$F(u) \geq -mu^2 - C_m ; \tag{14.34}$$

$$f(u)u - \varsigma F(u) + mu^2 \geq -C_m ; \tag{14.35}$$

$$|f'_u(u)| \leq C_0(1 + |u|^\rho) , \tag{14.36}$$

where $F(u) = \int_0^u f(v) dv$, $\lambda_1 > m > 0$, $\varsigma > 0$, C_m and C_0 are some constants, $2/(n-2) > \rho > 0$ for $n \geq 3$, and $\rho > 0$ is an arbitrary positive number for $n = 1, 2$.

The important examples of (14.33) give the nonlinearities

$$f(u) = \beta \sin u$$

and

$$f(u) = \beta |u|^\rho u .$$

In order to present more explicit expressions for dimension estimates let us restrict ourselves by the *sin-Gordon equation* with $f(u) = \beta \sin u$. More general consideration can be found elsewhere (see, for example, [318, 335]).

Let us introduce the following dimensionless numbers:

$$J = \frac{\beta^2}{\lambda_1 \gamma^2}, \quad D = \frac{\gamma^2}{\lambda_1} \tag{14.37}$$

For the sin-Gordon equation the dimension estimate of the attractor is $D_H(X) \leq m$, where the number m (see [318], p. 364) is the first integer such that

$$m \geq 2^7 \lambda_1 \sum_{i=1}^m \lambda_i^{-1} J(1 + D^2) . \tag{14.38}$$

For large i , $\lambda_i \sim \text{const} \lambda_1 i^{2/n}$, hence, for all i , $\lambda_1/\lambda_i < \text{const} i^{-2/n}$,

$$\lambda_1 \sum_{i=1}^m \lambda_i^{-1} < C m^{1-\frac{2}{n}} \tag{14.39}$$

for $n > 2$,

$$\lambda_1 \sum_{i=1}^m \lambda_i^{-1} < C \ln m \tag{14.40}$$

for $n = 2$, and

$$\lambda_1 \sum_{i=1}^m \lambda_i^{-1} < C \tag{14.41}$$

for $n = 1$. Here, in (14.39)–(14.40) constants depend only on the shape of Ω . Following these inequalities we can take as the upper dimension estimate the first integer m such that

$$m \geq (2^7 C J(1 + D^2))^{n/2} \tag{14.42}$$

for $n > 2$,

$$m \geq 2^7 C J(1 + D^2) \ln m \tag{14.43}$$

for $n = 2$, and

$$m \geq 2^7 C J(1 + D^2) \tag{14.44}$$

for $n = 1$.

The estimates for attractor dimension were also obtained for some other systems: thermohydraulics (including the Boussinesq equation), magnetohydrodynamics (see [318]), the Smoluchowski equation arising in the modeling of nematic liquid crystalline polymers [340], and others. Various dimensions (not only the Hausdorff dimension) were estimated.

14.4 Systems with Inheritance: Dynamics of Distributions with Conservation of Support, Natural Selection and Finite-Dimensional Asymptotics

14.4.1 Introduction: Unusual Conservation Law

In the 1970th-1980th years, theoretical studies developed one more “common” field belonging simultaneously to physics, biology and mathematics. For physics it is (so far) a part of the theory of approximations of a special kind, demonstrating, in particular, interesting mechanisms of discreteness in the course of evolution of distributions with initially smooth densities. But what for physics is merely a convenient approximation, is a fundamental law in biology (inheritance), whose consequences comprehended informally (selection theory [115, 377–380, 383, 384])² permeate most of the sections of this science.

Consider a community of animals. Let it be biologically isolated. Mutations can be neglected in the first approximation. In this case new genes do not emerge.

And here is an example is from physics. Let waves with wave vectors k be excited in some system. Denote K a set of wave vectors k of excited waves. Let the waves interaction do not lead to generation of waves with new $k \notin K$. Such an approximation is applicable to a variety of situations. For the wave turbulence it was described in detail in [385, 386].

What is common in these examples is the evolution of a distribution with a support not increasing in time.

What does not increase must, as a rule, decrease, if the decrease is not prohibited. This naive thesis can be converted into rigorous theorems for the case under consideration [115]. The support is proved to decrease in the limit $t \rightarrow \infty$, if it was sufficiently large initially. Considered usually are such system, for which at finite times the distributions support conserves and decrease only in the limit $t \rightarrow \infty$. Conservation of the support usually results in the following effect: dynamics of initially infinite-dimensional system at $t \rightarrow \infty$ can be described by finite-dimensional systems.

² We do not try to review the scientific literature about the evolution, and mention here only the references that are especially important for our understanding of the selection theory and applications.

The simplest and most common in applications class of equations for which the distributions support does not grow in time, is constructed as follows. To each distribution μ is assigned a function k_μ by which the distribution can be multiplied. Written down is equation

$$\frac{d\mu}{dt} = k_\mu \times \mu . \quad (14.45)$$

The multiplier k_μ is called a *reproduction coefficient*. The right-hand side is the product of the function k_μ with the distribution μ , hence $d\mu/dt$ should be zero where μ is equal to zero, therefore the support of μ is conserved in time (over the finite times).

Let us remind the definition of the support. Each distribution on a compact space X is a continuous linear functional on the space of continuous functions $C(X)$ ³. The space $C(X)$ is the Banach space endowed with the norm

$$\|f\| = \max_{x \in X} |f(x)| . \quad (14.46)$$

Usually, when X is a bounded closed subset of a finite-dimensional space, we represent this functional as the integral

$$\mu[f] = \int \mu(x)f(x) dx ,$$

where $\mu(x)$ is the (generalized) density function of the distribution μ . The support of μ , $\text{supp}\mu$, is the smallest closed subset of X with the following property: if $f(x) = 0$ on $\text{supp}\mu$, then $\mu[f] = 0$, i.e. $\mu(x) = 0$ outside $\text{supp}\mu$.

Strictly speaking, the space on which μ is defined and the distribution class it belongs to, should be specified. One should also specify are properties of the mapping $\mu \mapsto k_\mu$ and answer the question of existence and uniqueness of solutions of (14.45) under given initial conditions. In specific situations the answers to these questions are not difficult.

Let us start with the simplest example

$$\frac{\partial \mu(x, t)}{\partial t} = \left[f_0(x) - \int_a^b f_1(x)\mu(x, t) dx \right] \mu(x, t) , \quad (14.47)$$

where the functions $f_0(x)$ and $f_1(x)$ are positive and continuous on the closed segment $[a, b]$. Let the function $f_0(x)$ reaches the global maximum on the segment $[a, b]$ at a single point x_0 . If $x_0 \in \text{supp}\mu(x, 0)$, then

$$\mu(x, t) \rightarrow \frac{f_0(x_0)}{f_1(x_0)} \delta(x - x_0), \text{ when } t \rightarrow \infty , \quad (14.48)$$

³ We follow the Bourbaki approach [393]: a measure is a continuous functional, an integral. The book [393] contains all the necessary notions and theorems (and much more material than we need here).

where $\delta(x - x_0)$ is the δ -function.

We use in the space of measures the *weak convergence*, i.e. the convergence of averages:

$$\mu_i \rightarrow \mu^* \text{ if and only if } \int \mu_i \varphi(x) dx \rightarrow \int \mu^* \varphi(x) dx \quad (14.49)$$

for all continuous functions $\varphi(x)$. This weak convergence of measures generates the *weak topology* on the space of measures (the weak topology of conjugated space).

If $f_0(x)$ has several global maxima, then the right-hand side of (14.48) can be a sum of a finite number of δ -functions. Here a natural question arises: is it worth to pay attention to such a possibility? Should not we deem improbable for $f_0(x)$ to have more than one global maximum? Indeed, such a case seems to be very unlikely to occur. More details about this are given below.

Equations in the form (14.45) allow the following biological interpretation: μ is the distribution of the number (or of a biomass, or of another extensive variable) over inherited units: species, varieties, supergenes, genes. Whatever is considered as the *inherited unit* depends on the context, on a specific problem. The value of $k_\mu(x)$ is the reproduction coefficient of the inherited unit x under given conditions. The notion of “given conditions” includes the distribution μ , the reproduction coefficient depends on μ . Equation (14.47) can be interpreted as follows: $f_0(x)$ is the specific birth-rate of the inherited unit x (below, for the sake of definiteness, x is a variety, following the spirit of the famous Darwin’s book [377]), the death rate for the representatives of all inherited units (varieties) is determined by one common factor depending on the density $\int_a^b f_1(x)\mu(x, t) dx$; $f_1(x)$ is the individual contribution of the variety x into this death-rate.

On the other hand, for systems of waves with a parametric interaction, $k_\mu(x)$ can be the amplification (decay) rate of the wave with the wave vector x .

The first step in the routine of a dynamical system investigation is the question about fixed points and their stability. And the first observation concerning the system (14.45) is that asymptotically stable can be only the steady-state distributions, whose support is discrete (i.e. the sums of δ -functions). This can be proved for all the consistent formalizations, and can be understood as follows.

Let the “total amount” (integral of $|\mu|$ over U) be less than $\varepsilon > 0$ but not equal to zero in some domain U . Substitute distribution μ by zero on U , the rest remains as it is. It is natural to consider this disturbance of μ as ε -small. However, if the dynamics is described by (14.45), there is no way back to the undisturbed distribution, because the support cannot increase. If the steady state distribution μ^* is asymptotically stable, then for some $\varepsilon > 0$ any ε -small perturbation of μ^* relaxes back to μ^* . This is possible only in the case if for any domain U the integral of $|\mu^*|$ over U is either 0 or greater

than ε . Hence, this asymptotically stable distribution μ^* is the sum of the finite number of δ -functions:

$$\mu^*(x) = \sum_{i=1}^q N_i \delta(x - x_i) \quad (14.50)$$

with $|N_i| > \varepsilon$ for all i .

So we have: the support of asymptotically stable distributions for the system (14.45) is always discrete. This simple observation has many strong generalizations to general ω -limit points, to equations for vector measures, etc.

Dynamic systems where the phase variable is a distribution μ , and the distribution support is the integral of motion, frequently occur both in physics and in biology. Because of their attractive properties they are frequently used as approximations: we try to find the “main part” of the system in the form (14.45), and represent the rest as a small perturbation of the main part.

In biology such an approximation is essentially all the classical genetics, and also the formal contents of the theory of natural selection. The initial diversity is “thinned out” in time, and the limit distribution supports are described by some extremal principles (principles of optimality).

Conservation of the support in equation (14.45) can be considered as inheritance, and, consequently, we call the system (14.45) and its nearest generalizations “systems with inheritance”. Traditional division of the process of transferring biological information into inheritance and mutations, small in any admissible sense, can be compared to the description according to the following pattern: system (14.45) (or its nearest generalizations) plus small disturbances. Beyond the limits of such a description, talking about inheritance loses the conventional sense.

The first study of the dynamics systems with inheritance was due to J.B.S. Haldane. He used the simplest examples, studied steady-state distributions, and obtained the extremal principle for them. His pioneering book “The Causes of Evolution” (1932) [378] gives the clear explanation of the connections between the inheritance (the conservation of distributions support) and the optimality of selected varieties.

Haldane’s work was followed by entirely independent series of works on the S-approximation in the spin wave theory and on the wave turbulence [385–387], which studied wave configurations in the approximation of “inherited” wave vector, and by “Synergetics” [392], where the “natural selections” of modes is one of the basic concepts.

At the same time, a series of works on biological kinetics was done (see, for example, [115, 381–383]). The studies addressed not only steady-states, but also common limit distributions [115, 382] and waves in the space of inherited units [381]. For the steady-states a new type of stability was described – the *stable realizability* (see below).

The purpose of this paper is to present general results of the theory of systems with inheritance: optimality principles for limit distributions, theorems about selection, and estimations of the limit diversity (estimates of number of points in the supports of the limit distributions), drift effect and drift equations. Some of these results were published in preprints in Russian [382] (and, partially, in the Russian book [115]).

14.4.2 Optimality Principle for Limit Diversity

Description of the limit behavior of a dynamical system does not necessarily reduce to enumerating stable fixed points and limit cycles. The possibility of stochastic oscillations is common knowledge, while the domains of structurally unstable (non coarse) systems discovered by S. Smale [394]⁴ have so far not been mastered in applied and natural sciences).

The leading rival to adequately formalize the limit behavior is the concept of the “ ω -limit set”. It was discussed in detail in the classical monograph [395]. The fundamental textbook on dynamical systems [396] and the introductory review [397] are also available.

If $f(t)$ is the dependence of the position of point in the phase space on time t (i.e. the *motion* of the dynamical system), then the ω -limit points are such points y , for which there exist such sequences of times $t_i \rightarrow \infty$, that $f(t_i) \rightarrow y$.

The set of all ω -limit points for the given motion $f(t)$ is called the ω -limit set. If, for example, $f(t)$ tends to the equilibrium point y^* then the corresponding ω -limit set consists of this equilibrium point. If $f(t)$ is winding onto a closed trajectory (the limit cycle), then the corresponding ω -limit set consists of the points of the cycle and so on.

General ω -limit sets are not encountered often in specific situations. This is because of the lack of efficient methods to find them in a general situation. Systems with inheritance is a case, where there are efficient methods to estimate the limit sets from above. This is done by the optimality principle.

Let $\mu(t)$ be a solution of (14.45). Note that

$$\mu(t) = \mu(0) \exp \int_0^t k_{\mu(\tau)} d\tau . \quad (14.51)$$

⁴ “Structurally stable systems are not dense”. Without exaggeration we can say that so entitled work [394] opened a new era in the understanding of dynamics. Structurally stable (rough) systems are those whose phase portraits do not change qualitatively under small perturbations. Smale constructed such structurally unstable system that any other system close enough to it is also structurally unstable. This result defeated hopes for a classification if not all, but at least “almost all” dynamical systems. Such hopes were associated with the success of the classification of two-dimensional dynamical systems, among which structurally stable systems are dense.

Here and below we do not display the dependence of distributions μ and of the reproduction coefficients k on x when it is not necessary. Fix the notation for the average value of $k_{\mu(\tau)}$ on the segment $[0, t]$

$$\langle k_{\mu(t)} \rangle_t = \frac{1}{t} \int_0^t k_{\mu(\tau)} \, d\tau . \tag{14.52}$$

Then the expression (14.51) can be rewritten as

$$\mu(t) = \mu(0) \exp(t \langle k_{\mu(t)} \rangle_t) . \tag{14.53}$$

If μ^* is the ω -limit point of the solution $\mu(t)$, then there exists such a sequence of times $t_i \rightarrow \infty$, that $\mu(t_i) \rightarrow \mu^*$. Let it be possible to chose a convergent subsequence of the sequence of the average reproduction coefficients $\langle k_{\mu(t)} \rangle_t$, which corresponds to times t_i . We denote as k^* the limit of this subsequence. Then, the following statement is valid: on the support of μ^* the function k^* vanishes and on the support of $\mu(0)$ it is non-positive:

$$\begin{aligned} k^*(x) &= 0 \text{ if } x \in \text{supp}\mu^* , \\ k^*(x) &\leq 0 \text{ if } x \in \text{supp}\mu(0) . \end{aligned} \tag{14.54}$$

Taking into account the fact that $\text{supp}\mu^* \subseteq \text{supp}\mu(0)$, we come to the formulation of **the optimality principle** (14.54): *The support of limit distribution consists of points of the global maximum of the average reproduction coefficient on the initial distribution support. The corresponding maximum value is zero.*

We should also note that not necessarily all points of maximum of k^* on $\text{supp}\mu(0)$ belong to $\text{supp}\mu^*$, but all points of $\text{supp}\mu^*$ are the points of maximum of k^* on $\text{supp}\mu(0)$.

If $\mu(t)$ tends to the fixed point μ^* , then $\langle k_{\mu(t)} \rangle_t \rightarrow k_{\mu^*}$ as $t \rightarrow \infty$, and $\text{supp}\mu^*$ consists of the points of the global maximum of the corresponding reproduction coefficient k_{μ^*} on the support of μ^* . The corresponding maximum value is zero.

If $\mu(t)$ tends to the limit cycle $\mu^*(t)$ ($\mu^*(t+T) = \mu^*(t)$), then all the distributions $\mu^*(t)$ have the same support. The points of this support are the points of maximum (global, zero) of the averaged over the cycle reproduction coefficient

$$k^* = \langle k_{\mu^*(t)} \rangle_T = \frac{1}{T} \int_0^T k_{\mu^*(\tau)} \, d\tau , \tag{14.55}$$

on the support of $\mu(0)$.

The supports of the ω -limit distributions are specified by the functions k^* . It is obvious where to get these functions from for the cases of fixed points and limit cycles. There are at least two questions: what ensures the existence of average reproduction coefficients at $t \rightarrow \infty$, and how to use the described extremal principle (and how efficient is it). The latter question is

the subject to be considered in the following sections. In the situation to follow the answers to these questions have the validity of theorems. Let X be a space on which the distributions are defined. Assume it to be a compact metric space (for example, a closed bounded subset of Euclidian space). The distribution μ is identified with the Radon measure, that is the continuous linear functional on the space of continuous functions on X , $C(X)$. We use the conventional notation for this linear functional as the integral of the function φ as $\int \varphi(x)\mu(x) dx$. Here $\mu(x)$ is acting as the distribution density, although, of course, the arbitrary X has no initial (Lebesgue, for example) dx .

The sequence of continuous functions $k_i(x)$ is considered to be convergent if it converges uniformly. The sequence of measures μ_i is called convergent if for any continuous function $\varphi(x)$ the integrals $\int \varphi(x)\mu_i(x) dx$ converge (weak convergence (14.49)). The mapping $\mu \mapsto k_\mu$ assigning the reproduction coefficient k_μ to the measure μ is assumed to be continuous. And, finally, the space of measures is assumed to have a bounded⁵ set M which is positively invariant relative to system (14.45): if $\mu(0) \in M$, then $\mu(t) \in M$ (and is non-trivial). This M will serve as the phase space of system (14.45).

Most of the results about systems with inheritance use **the theorem about weak compactness**: *The bounded set of measures is precompact with respect to the weak convergence (i.e., its closure is compact)*. Therefore, the set of corresponding reproduction coefficients $k_M = \{k_\mu | \mu \in M\}$ is precompact, the set of averages (14.52) is precompact, because it is the subset of the closed convex hull $\overline{\text{conv}}(k_M)$ of the compact set. This compactness allows us to claim the existence of the *average reproduction coefficient* k^* for the description of the ω -limit distribution μ^* with the optimality principle (14.54).

14.4.3 How Many Points Does the Limit Distribution Support Hold?

The limit distribution is concentrated in the points of (zero) global maximum of the average reproduction coefficient. The average is taken along the solution, but the solution is not known beforehand. With the convergence towards a fixed point or to a limit cycle this difficulty can be circumvented. In the general case the extremal principle can be used without knowing the solution, in the following way [115]. Considered is a set of all dependencies $\mu(t)$ where μ belongs to the phase space, the bounded set M . The set of all averages over t is $\{ \langle k_{\mu(t)} \rangle_t \}$. Further, taken are all limits of sequences formed by these averages – the set of averages is closed. The result is the closed convex hull $\overline{\text{conv}}(k_M)$ of the compact set k_M . This set involves all possible averages (14.52) and all their limits. In order to construct it, the true solution $\mu(t)$ is not needed.

⁵ The set of measures M is bounded, if the sets of integrals $\{\mu[f] | \mu \in M, \|f\| \leq 1\}$ is bounded, where $\|f\|$ is the norm (14.46).

The *weak optimality principle* is expressed as follows. Let $\mu(t)$ be a solution of (14.45) in M , μ^* is any of its ω -limit distributions. Then in the set $\overline{\text{conv}}(k_M)$ there is such a function k^* that its maximum value on the support $\text{supp}\mu_0$ of the initial distribution μ_0 equals to zero, and $\text{supp}\mu^*$ consists of the points of the global maximum of k^* on $\text{supp}\mu_0$ only (14.54).

Of course, in the set $\overline{\text{conv}}(k_M)$ usually there are many functions that are irrelevant to the time average reproduction coefficients for the given motion $\mu(t)$. Therefore, the weak extremal principle is really weak – it gives too many possible supports of μ^* . However, even such a principle can help to obtain useful estimates of the number of points in the supports of ω -limit distributions.

It is not difficult to suggest systems of the form (14.45), in which any set can be the limit distribution support. The simplest example: $k_\mu \equiv 0$. Here ω -limit (fixed) is any distribution. However, almost any arbitrary small perturbation of the system destroys this pathological property.

In the realistic systems, especially in biology, the coefficients fluctuate and are never known exactly. Moreover, the models are in advance known to have a finite error which cannot be exterminated by the choice of the parameters values. This gives rise to an idea to consider not individual systems (14.45), but ensembles of similar systems [115].

Having posed the questions of how many points can the support of ω -limit distributions have, estimate the maximum for each individual system from the ensemble (in its ω -limit distributions), and then, estimate the minimum of these maxima over the whole ensemble – (*the minimax estimation*). The latter is motivated by the fact, that if the inherited unit has gone extinct under some conditions, it will not appear even under the change of conditions.

Let us consider an ensemble that is simply the ε -neighborhood of the given system (14.45). The minimax estimates of the number of points in the support of ω -limit distribution are constructed by approximating the dependencies k_μ by finite sums

$$k_\mu = \varphi_0(x) + \sum_{i=1}^n \varphi_i(x)\psi_i(\mu). \tag{14.56}$$

Here φ_i depend on x only, and ψ_i depend on μ only. Let $\varepsilon_n > 0$ be the distance from k_μ to the nearest sum (14.56) (the “distance” is understood in the suitable rigorous sense, which depends on the specific problem). So, we reduced the problem to the estimation of the diameters $\varepsilon_n > 0$ of the set $\overline{\text{conv}}(k_M)$.

The minimax estimation of the number of points in the limit distribution support gives the answer to the question, “How many points does the limit distribution support hold”: *If $\varepsilon > \varepsilon_n$ then, in the ε -vicinity of k_μ , the minimum of the maxima of the number of points in the ω -limit distribution support does not exceed n .*

In order to understand this estimate it is sufficient to consider system (14.45) with k_μ of the form (14.56). The averages (14.52) for any dependence

$\mu(t)$ in this case have the form

$$\langle k_{\mu(t)} \rangle_t = \frac{1}{t} \int_0^t k_{\mu(\tau)} d\tau = \varphi_0(x) + \sum_{i=1}^n \varphi_i(x) a_i . \quad (14.57)$$

where a_i are some numbers. The ensemble of the functions (14.57) for various a_i forms a n -dimensional linear manifold. How many of points of the global maximum (equal to zero) could a function of this family have?

Generally speaking, it can have any number of maxima. However, it seems obvious, that “usually” one function has only one point of global maximum, while it is “improbable” the maximum value is zero. At least, with an arbitrary small perturbation of the given function, we can achieve for the point of the global maximum to be unique and the maximum value be non-zero.

In a one-parametric family of functions there may occur zero value of the global maximum, which cannot be eliminated by a small perturbation, and individual functions of the family may have two global maxima.

In the general case we can state, that “usually” each function of the n -parametric family (14.57) can have not more than $n - 1$ points of the zero global maximum (of course, there may be less, and for the majority of functions of the family the global maximum, as a rule, is not equal to zero at all). What “usually” means here requires a special explanation given in the next section.

In application k_μ is often represented by an integral operator, linear or nonlinear. In this case the form (14.56) corresponds to the kernels of integral operators, represented in a form of the sums of functions' products. For example, the reproduction coefficient of the following form

$$k_\mu = \varphi_0(x) + \int K(x, y) \mu(y) dy ,$$

$$\text{where } K(x, y) = \sum_{i=1}^n \varphi_i(x) g_i(y) , \quad (14.58)$$

has also the form (14.56) with $\psi_i(\mu) = \int g_i(y) \mu(y) dy$.

The linear reproduction coefficients occur in applications rather frequently. For them the problem of the minimax estimation of the number of points in the ω -limit distribution support is reduced to the question of the accuracy of approximation of the linear integral operator by the sums of kernels-products (14.58).

14.4.4 Selection Efficiency

The first application of the extremal principle for the ω -limit sets is the theorem of the selection efficiency. The dynamics of a system with inheritance indeed leads in the limit $t \rightarrow \infty$ to a selection. In the typical situation, a

diversity in the limit $t \rightarrow \infty$ becomes less than the initial diversity. There is an efficient selection for the “best”. The basic effects of selection are formulated below.

Theorem of Selection Efficiency

1. *Almost always the support of any ω -limit distribution is nowhere dense in X (and it has the Lebesgue measure zero for Euclidean space).*
2. *Let $\varepsilon_n > 0$, $\varepsilon_n \rightarrow 0$ be an arbitrary chosen sequence. The following statement is almost always true for system (14.45). Let the support of the initial distribution be the whole X . Then the support of any ω -limit distribution μ^* is almost finite. This means that it is approximated by finite sets faster than $\varepsilon_n \rightarrow 0$: for any $\delta > 0$ there is such a number N that for any ω -limit distribution μ^* there exists a finite set S_N of N elements such that $\text{dist}(S_N, \text{supp}\mu^*) < \delta\varepsilon_N$, where dist is the Hausdorff distance:*

$$\text{dist}(A, B) = \max\{\sup_{x \in A} \inf_{x \in B} \rho(x, y), \sup_{x \in B} \inf_{x \in A} \rho(x, y)\}, \tag{14.59}$$

where $\rho(x, y)$ is the distance between points.

3. *In the previous statement for any chosen sequence $\varepsilon_n > 0$, $\varepsilon_n \rightarrow 0$, almost all systems (14.45) have ω -limit distributions with supports that can be approximated by finite sets faster than $\varepsilon_n \rightarrow 0$. The order is important: “for any sequence almost all systems...” But if we use only the recursive (algorithmic) analogue of sequences, then we can easily prove the statement with the reverse order: “almost all systems for any sequence...” This is possible because the set of all recursive enumerable countable sets is also countable and not continuum. This observation is very important for algorithmic foundations of probability theory [398]. Let \mathbf{L} be a set of all sequences of real numbers $\varepsilon_n > 0$, $\varepsilon_n \rightarrow 0$ with the property: for each $\{\varepsilon_n\} \in \mathbf{L}$ the rational subgraph $\{(n, r) : \varepsilon_n > r \in \mathbf{Q}\}$ is recursively enumerable. For almost all systems (14.45) and any $\{\varepsilon_n\} \in \mathbf{L}$ the support of any ω -limit distribution μ^* is approximated by finite sets faster than $\varepsilon_n \rightarrow 0$.*

These properties hold for the continuous reproduction coefficients. It is well-known, that it is dangerous to rely on the genericity among continuous functions. For example, almost all continuous functions are nowhere differentiable. But the properties 1 and 2 hold also for the smooth reproduction coefficients on the manifolds and sometimes allow to replace the “almost finiteness” by simply finiteness. In order to appreciate this theorem, note that:

1. Support of an arbitrary ω -limit distribution μ^* consist of points of global maximum of the average reproduction coefficient on a support of the initial distribution. The corresponding maximum value is zero.
2. Almost always a function has only one point of global maximum, and corresponding maximum value is not 0.

3. In a one-parametric family of functions almost always there may occur zero values of the global maximum (at one point), which cannot be eliminated by a small perturbation, and individual functions of the family may stably have two global maximum points.
4. For a generic n -parameter family of functions, there may exist stably a function with $n - 1$ points of global maximum and with zero value of this maximum.
5. Our phase space M is compact. The set of corresponding reproduction coefficients k_M in $C(X)$ for the given map $\mu \rightarrow k_\mu$ is compact too. The average reproduction coefficients belong to the closed convex hull of this set $\overline{\text{conv}}(k_M)$. And it is compact too.
6. A compact set in a Banach space can be approximated by its projection on an appropriate finite-dimensional linear manifold with an arbitrary accuracy. Almost always the function on such a manifold may have only $n - 1$ points of global maximum with zero value, where n is the dimension of the manifold.

The rest of the proof is purely technical. The easiest demonstration of the “natural” character of these properties is the demonstration of instability of exclusions: If, for example, a function has several points of global maxima then with an arbitrary small perturbation (for all usually used norms) it can be transformed into a function with the unique point of global maximum. However “stable” does not always mean “dense”. In what sense the discussed properties of the system (14.45) are usually valid? “Almost always”, “typically”, “generically” a function has only one point of global maximum. This sentence should be given an rigorous meaning. Formally it is not difficult, but haste is dangerous when defining “genericity”.

Here are some examples of correct but useless statements about “generic” properties of function: Almost every continuous function is not differentiable; Almost every C^1 -function is not convex. Their meaning for applications is most probably this: the genericity used above for continuous functions or for C^1 -function is irrelevant to the subject.

Most frequently the motivation for definitions of genericity is found in such a situation: given n equations with m unknowns, what can we say about the solutions? The answer is: in a typical situation, if there are more equations, than the unknowns ($n > m$), there are no solutions at all, but if $n \leq m$ (n is less or equal to m), then, either there is a $(m - n)$ -parametric family of solutions, or there are no solutions.

The best known example of using this reasoning is the *Gibbs phase rule* in classical chemical thermodynamics. It limits the number of co-existing phases. There exists a well-known example of such reasoning in mathematical biophysics too. Let us consider a medium where n species coexist. The medium is assumed to be described by m parameters. In the simplest case, the medium is a well-mixed solution of m substances. Let the organisms interact through the medium, changing its parameters – concentrations of m

substances. Then, in a steady state, for each of the coexisting species we have an equation with respect to the state of the medium. So, the number of such species cannot exceed the number of parameters of the medium. In a typical situation, in the m -parametric medium in a steady state there can exist not more than m species. This is the *Gause concurrent exclusion principle* [399]. This fact allows numerous generalizations. Theorem of the natural selection efficiency may be considered as its generalization too.

Analogous assertion for a non-steady state coexistence of species in the case of equations (11) is not true. It is not difficult to give an example of stable coexistence under oscillating conditions of n species in the m -parametric medium at $n > m$. But, if k_μ are linear functions of μ , then for non-stable conditions we have the concurrent exclusion principle, too. In that case, the average in time of reproduction coefficient $k_{\mu(t)}$ is the reproduction coefficient for the average $\mu(t)$ because of linearity. Therefore, the equation for the average reproduction coefficient,

$$k^*(x) = 0 \text{ for } x \in \text{supp}\mu^* , \quad (14.60)$$

transforms into the following equation for the reproduction coefficient of the average distribution

$$k^*(\langle\mu\rangle) = 0 \text{ for } x \in \text{supp}\mu^* \quad (14.61)$$

(the *Volterra averaging principle* [400]). This system has as many linear equations as it has coexisting species. The averages can be non-unique. Then all of them satisfy this system, and we obtain the non-stationary Gause principle. And again, it is valid “almost always”.

Formally, various definitions of genericity are constructed as follows. All systems (or cases, or situations and so on) under consideration are somehow parameterized – by sets of vectors, functions, matrices etc. Thus, the “space of systems” Q can be described. Then the “thin sets” are introduced into Q , i.e. the sets, which we shall later neglect. The union of a finite or countable number of thin sets, as well as the intersection of any number of them should be thin again, while the whole Q is not thin. There are two traditional ways to determine thinness.

1. A set is considered thin when it has measure zero. This is good for a finite-dimensional case, when there is the standard Lebesgue measure – the length, the area, the volume.
2. But most frequently we deal with the functional parameters. In that case it is common to restore to the second definition, according to which the sets of first category are negligible. The construction begins with nowhere dense sets. The set Y is nowhere dense in Q , if in any nonempty open set $V \subset Q$ (for example, in a ball) there exists a nonempty open subset $W \subset V$ (for example, a ball), which does not intersect with Y . Roughly speaking, Y is “full of holes” – in any neighborhood of any point of the set Y there is an open hole. Countable union of nowhere dense sets is called

the set of the first category. The second usual way is to define thin sets as the *sets of the first category*.

But even the real line R can be divided into two sets, one of which has zero measure, the other is of the first category. The genericity in the sense of measure and the genericity in the sense of category considerably differ in the applications where both of these concepts can be used. The conflict between the two main views on genericity stimulated efforts to invent new and stronger approaches.

In Theorem of selection efficiency a very strong genericity was used. Systems (14.45) were parameterized by continuous maps $\mu \mapsto k_\mu$. Denote by Q the space of these maps $M \rightarrow C(X)$ with the topology of uniform convergence on M . So, it is a Banach space. We shall call the set Y in the Banach space Q *completely thin*, if for any compact set K in Q and arbitrary positive $\varepsilon > 0$ there exists a vector $q \in Q$, such that $\|q\| < \varepsilon$ and $K + q$ does not intersect Y . So, a set, which can be moved out of intersection with any compact by an arbitrary small translation, is *completely negligible*. In a finite-dimensional space there is only one such set – the empty one. In an infinite-dimensional Banach space compacts and closed subspaces with infinite codimension provide us examples of completely negligible sets. In *Theorem of selection efficiency* “usually” means “the set of exceptions is completely thin”.

14.4.5 Gromov’s Interpretation of Selection Theorems

In his talk [401], M. Gromov offered a geometric interpretation of the selection theorems. Let us consider dynamical systems in the standard m -simplex σ_m in $m + 1$ -dimensional space R^{m+1} :

$$\sigma_m = \{x \in R^{m+1} | x_i \geq 0, \sum_{i=1}^{m+1} x_i = 1\}.$$

We assume that simplex σ_m is positively invariant with respect to these dynamical systems: if the motion starts in σ_m at some time t_0 then it remains in σ_m for $t > t_0$. Let us consider the motions that start in the simplex σ_m at $t = 0$ and are defined for $t > 0$.

For large m , almost all volume of the simplex σ_m is concentrated in a small neighborhood of the center of σ_m , near the point $c = (\frac{1}{m}, \frac{1}{m}, \dots, \frac{1}{m})$. Hence, one can expect that a typical motion of a general dynamical system in σ_m for sufficiently large m spends almost all the time in a small neighborhood of c .

Let us consider dynamical systems with an additional property (“inheritance”): all the faces of the simplex σ_m are also positively invariant with respect to the systems with inheritance. It means that if some $x_i = 0$ initially at the time $t = 0$ then $x_i = 0$ for $t > 0$ for all motions in σ_m . The essence of selection theorems is as follows: a typical motion of a typical dynamical

system with inheritance spends almost all the time in a small neighborhood of low-dimensional faces, even if it starts near the center of the simplex.

Let us denote by $\partial_r \sigma_m$ the union of all r -dimensional faces of σ_m . Due to the selection theorems, a typical motion of a typical dynamical system with inheritance spends almost all time in a small neighborhood of $\partial_r \sigma_m$ with $r \ll m$. It should not obligatory reside near just one face from $\partial_r \sigma_m$, but can travel in neighborhood of different faces from $\partial_r \sigma_m$ (the drift effect). The minimax estimation of the number of points in ω -limit distributions through the diameters $\varepsilon_n > 0$ of the set $\overline{\text{conv}}(k_M)$ is the estimation of r .

14.4.6 Drift Equations

To this end, we talked about the support of an individual ω -limit distribution. Almost always it is small. But this does not mean, that the union of these supports is small even for one solution $\mu(t)$. It is possible that a solution is a finite set of narrow peaks getting in time more and more narrow, moving slower and slower, but not tending to fixed positions, rather continuing to move along its trajectory, and the path covered tends to infinity as $t \rightarrow \infty$.

This effect was not discovered for a long time because the slowing down of the peaks was thought as their tendency to fixed positions. There are other difficulties related to the typical properties of continuous functions, which are not typical for the smooth ones. Let us illustrate them for the distributions over a straight line segment. Add to the reproduction coefficients k_μ the sum of small and narrow peaks located on a straight line distant from each other much more than the peak width (although it is ε -small). However small is chosen the peak's height, one can choose their width and frequency on the straight line in such a way that from any initial distribution μ_0 whose support is the whole segment, at $t \rightarrow \infty$ we obtain ω -limit distributions, concentrated at the points of maximum of the added peaks.

Such a model perturbation is small in the space of continuous functions. Therefore, it can be put as follows: *by small continuous perturbation the limit behavior of system (14.45) can be reduced onto a ε -net for sufficiently small ε* . But this can not be done with the small smooth perturbations (with small values of the first and the second derivatives) in the general case. The discreteness of the net, onto which the limit behavior is reduced by small continuous perturbations, differs from the discreteness of the support of the individual ω -limit distribution. For an individual distribution the number of points is estimated, roughly speaking, by the number of essential parameters (14.56), while for the conjunction of limit supports – by the number of stages in approximation of k_μ by piece-wise constant functions.

Thus, in a typical case the dynamics of systems (14.45) with smooth reproduction coefficients transforms a smooth initial distributions into the ensemble of narrow peaks. The peaks become more narrow, their motion slows down, but not always they tend to fixed positions.

The equations of motion for these peaks can be obtained in the following way [115]. Let X be a domain in the n -dimensional real space, and the initial distributions μ_0 be assumed to have smooth density. Then, after sufficiently large time t , the position of distribution peaks are the points of the average reproduction coefficient maximum $\langle k_\mu \rangle_t$ (14.52) to any accuracy set in advance. Let these points of maximum be x^α , and

$$q_{ij}^\alpha = -t \frac{\partial^2 \langle k_\mu \rangle_t}{\partial x_i \partial x_j} \Big|_{x=x^\alpha} .$$

It is easy to derive the following differential relations

$$\begin{aligned} \sum_j q_{ij}^\alpha \frac{dx_j^\alpha}{dt} &= \frac{\partial k_{\mu(t)}}{\partial x_i} \Big|_{x=x^\alpha} ; \\ \frac{dq_{ij}^\alpha}{dt} &= - \frac{\partial^2 k_{\mu(t)}}{\partial x_i \partial x_j} \Big|_{x=x^\alpha} . \end{aligned} \tag{14.62}$$

These relations do not form a closed system of equations, because the right-hand parts are not functions of x_i^α and q_{ij}^α . For sufficiently narrow peaks there should be separation of the relaxation times between the dynamics *on* the support and the dynamics *of* the support: the relaxation of peak amplitudes (it can be approximated by the relaxation of the distribution with the finite support, $\{x^\alpha\}$) should be significantly faster than the motion of the locations of the peaks, the dynamics of $\{x^\alpha\}$. Let us write the first term of the corresponding asymptotics [115].

For the finite support $\{x^\alpha\}$ the distribution is $\mu = \sum_\alpha N_\alpha \delta(x - x^\alpha)$. Dynamics of the finite number of variables, N_α obeys the system of ordinary differential equations

$$\frac{dN_\alpha}{dt} = k_\alpha(\mathbf{N}) N_\alpha , \tag{14.63}$$

where \mathbf{N} is vector with components N_α , $k_\alpha(\mathbf{N})$ is the value of the reproduction coefficient k_μ at the point x^α :

$$k_\alpha(\mathbf{N}) = k_\mu(x^\alpha) \text{ for } \mu = \sum_\alpha N_\alpha \delta(x - x^\alpha) .$$

Let the dynamics of the system (14.63) for a given set of initial conditions be simple: the motion $\mathbf{N}(t)$ goes to the stable fixed point $\mathbf{N} = \mathbf{N}^*({x^\alpha})$. Then we can take in the right hand side of (14.62)

$$\mu(t) = \mu^*({x^\alpha(t)}) = \sum_\alpha N_\alpha^* \delta(x - x^\alpha(t)) . \tag{14.64}$$

Because of the time separation we can assume that (i) relaxation of the amplitudes of peaks is completed and (ii) peaks are sufficiently narrow, hence, the difference between true $k_{\mu(t)}$ and the reproduction coefficient for the

measure (14.64) with the finite support $\{x^\alpha\}$ is negligible. Let us use the notation $k^*(\{x^\alpha\})(x)$ for this reproduction coefficient. The relations (14.62) transform into the ordinary differential equations

$$\begin{aligned} \sum_j q_{ij}^\alpha \frac{dx_j^\alpha}{dt} &= \frac{\partial k^*(\{x^\beta\})(x)}{\partial x_i} \Big|_{x=x^\alpha} ; \\ \frac{dq_{ij}^\alpha}{dt} &= - \frac{\partial^2 k^*(\{x^\beta\})(x)}{\partial x_i \partial x_j} \Big|_{x=x^\alpha} . \end{aligned} \quad (14.65)$$

For many purposes it may be useful to switch to the logarithmic time $\tau = \ln t$ and to new variables

$$b_{ij}^\alpha = \frac{1}{t} q_{ij}^\alpha = - \frac{\partial^2 \langle k(\mu) \rangle_t}{\partial x_i \partial x_j} \Big|_{x=x^\alpha} .$$

For large t we obtain from (14.65)

$$\begin{aligned} \sum_j b_{ij}^\alpha \frac{dx_j^\alpha}{d\tau} &= \frac{\partial k^*(\{x^\beta\})(x)}{\partial x_i} \Big|_{x=x^\alpha} ; \\ \frac{db_{ij}^\alpha}{d\tau} &= - \frac{\partial^2 k^*(\{x^\alpha\})(x)}{\partial x_i \partial x_j} \Big|_{x=x^\beta} - b_{ij}^\alpha . \end{aligned} \quad (14.66)$$

The way of constructing the drift equations (14.65,14.66) for a specific system (14.45) is as follows:

1. For finite sets $\{x^\alpha\}$ one studies systems (14.63) and finds the equilibrium solutions $N^*(\{x^\alpha\})$;
2. For given measures $\mu^*(\{x^\alpha(t)\})$ (14.64) one calculates the reproduction coefficients $k_\mu(x) = k^*(\{x^\alpha\})(x)$ and first derivatives of these functions in x at points x^α . That is all, the drift equations (14.65,14.66) are set up.

The drift equations (14.65,14.66) describe the dynamics of the peaks positions x^α and of the coefficients q_{ij}^α . For given x^α , q_{ij}^α , and N_α^* the distribution density μ can be approximated as the sum of narrow Gaussian peaks:

$$\mu = \sum_\alpha N_\alpha^* \sqrt{\frac{\det Q^\alpha}{(2\pi)^n}} \exp \left(- \frac{1}{2} \sum_{ij} q_{ij}^\alpha (x_i - x_i^\alpha)(x_j - x_j^\alpha) \right) , \quad (14.67)$$

where Q^α is the inverse covariance matrix (q_{ij}^α) .

If the limit dynamics of the system (14.63) for finite supports at $t \rightarrow \infty$ can be described by a more complicated attractor, then instead of reproduction coefficient $k^*(\{x^\alpha\})(x) = k_{\mu^*}$ for the stationary measures μ^* (14.64) one can use the average reproduction coefficient with respect to the corresponding *Sinai–Ruelle–Bowen measure* [396,397]. If finite systems (14.63) have several attractors for given $\{x^\alpha\}$, then the dependence $k^*(\{x^\alpha\})$ is multi-valued, and

there may be bifurcations and hysteresis with the function $k^*({x^\alpha})$ transition from one sheet to another. There are many interesting effects concerning peaks' birth, desintegration, divergence, and death, and the drift equations (14.65,14.66) describe the motion in a non-critical domain, between these critical effects.

Inheritance (conservation of support) is never absolutely exact. Small variations, mutations, immigration in biological systems are very important. Excitation of new degrees of freedom, modes diffusion, noise are present in physical systems. How does small perturbation in the inheritance affect the effects of selection? The answer is usually as follows: there is such a value of perturbation of the right-hand side of (14.45), at which they would change nearly nothing, just the limit δ -shaped peaks transform into sufficiently narrow peaks, and zero limit of the velocity of their drift at $t \rightarrow \infty$ substitutes by a small finite one.

The simplest model for "inheritance + small variability" is given by a perturbation of (14.45) with diffusion term

$$\frac{\partial \mu(x, t)}{\partial t} = k_{\mu(x, t)} \times \mu(x, t) + \varepsilon \sum_{ij} d_{ij}(x) \frac{\partial^2 \mu(x, t)}{\partial x_i \partial x_j} . \tag{14.68}$$

where $\varepsilon > 0$ and the matrix of diffusion coefficients d_{ij} is symmetric and positively definite.

There are almost always no qualitative changes in the asymptotic behaviour, if ε is sufficiently small. With this the asymptotics is again described by the drift equations (14.65, 14.66), modified by taking into account the diffusion as follows:

$$\begin{aligned} \sum_j q_{ij}^\alpha \frac{dx_j^\alpha}{dt} &= \left. \frac{\partial k^*({x^\beta})(x)}{\partial x_i} \right|_{x=x^\alpha} ; \\ \frac{dq_{ij}^\alpha}{dt} &= - \left. \frac{\partial^2 k^*({x^\beta})(x)}{\partial x_i \partial x_j} \right|_{x=x^\alpha} - 2\varepsilon \sum_{kl} q_{ik}^\alpha d_{kl}(x^\alpha) q_{lj}^\alpha . \end{aligned} \tag{14.69}$$

Now, as distinct from (14.65), the eigenvalues of the matrices $Q^\alpha = (q_{ij}^\alpha)$ cannot grow infinitely. This is prevented by the quadratic terms in the right-hand side of the second equation (14.69).

Dynamics of (14.69) does not depend on the value $\varepsilon > 0$ qualitatively, because of the obvious scaling property. If ε is multiplied by a positive number ν , then, upon rescaling $t' = \nu^{-1/2}t$ and $q_{ij}^{\alpha'} = \nu^{-1/2}q_{ij}^\alpha$, we have the same system again. Multiplying $\varepsilon > 0$ by $\nu > 0$ changes only peak's velocity values by a factor $\nu^{1/2}$, and their width by a factor $\nu^{1/4}$. The paths of peaks' motion do not change at this for the drift approximation (14.69) (but the applicability of this approximation may, of course, change).

14.4.7 Three Main Types of Stability

Stable steady-state solutions of equations of the form (14.45) may be only the sums of δ -functions – this was already mentioned. There is a set of specific conditions of stability, determined by the form of equations.

Consider a stationary distribution for (14.45) with a finite support

$$\mu^*(x) = \sum_{\alpha} N_{\alpha}^* \delta(x - x^{*\alpha}) .$$

Steady state of μ^* means, that

$$k_{\mu^*}(x^{*\alpha}) = 0 \text{ for all } \alpha . \quad (14.70)$$

The *internal stability* means, that this distribution is stable with respect to perturbations not increasing the support of μ^* . That is, the vector N_{α}^* is the stable fixed point for the dynamical system (14.63). Here, as usual, it is possible to distinguish between the Lyapunov stability, the asymptotic stability and the first approximation stability (negativeness of real parts for the eigenvalues of the matrix $\partial \dot{N}_{\alpha}^* / \partial N_{\alpha}^*$ at the stationary points).

The *external stability* means stability to an expansion of the support, i.e. to adding to μ^* of a small distribution whose support contains points not belonging to $\text{supp} \mu^*$. It makes sense to speak about the external stability only if there is internal stability. In this case it is sufficient to restrict ourselves with δ -functional perturbations. The external stability has a very transparent physical and biological sense. It is stability with respect to *introduction* into the systems of a new inherited unit (gene, variety, specie...) in a small amount.

The *necessary condition for the external stability* is: the points $\{x^{*\alpha}\}$ are points of the global maximum of the reproduction coefficient $k_{\mu^*}(x)$. It can be formulated as the optimality principle

$$k_{\mu^*}(x) \leq 0 \text{ for all } x; \quad k_{\mu^*}(x^{*\alpha}) = 0 . \quad (14.71)$$

The *sufficient condition for the external stability* is: the points $\{x^{*\alpha}\}$ and only these points are points of the global maximum of the reproduction coefficient $k_{\mu^*}(x)$. At the same time it is the condition of the external stability in the first approximation and the optimality principle

$$k_{\mu^*}(x) < 0 \text{ for } x \notin \{x^{*\alpha}\}; \quad k_{\mu^*}(x^{*\alpha}) = 0 . \quad (14.72)$$

The only difference from (14.71) is the change of the inequality sign from $k_{\mu^*}(x) \leq 0$ to $k_{\mu^*}(x) < 0$ for $x \notin \{x^{*\alpha}\}$. The necessary condition (14.71) means, that the small δ -functional addition will not grow in the first approximation. According to the sufficient condition (14.72) such a small addition will exponentially decrease.

If X is a finite set then the combination of the external and the internal stability is equivalent to the standard stability for a system of ordinary differential equations.

For the continuous X there is one more kind of stability important from the applications viewpoint. Substitute δ -shaped peaks at the points $\{x^{*\alpha}\}$ by narrow Gaussians and shift slightly the positions of their maxima away from the points $x^{*\alpha}$. How will the distribution from such initial conditions evolve? If it tends to μ without getting too distant from this steady state distribution, then we can say that the third type of stability – *stable realizability* – takes place. It is worth mentioning that the perturbation of this type is only weakly small, in contrast to perturbations considered in the theory of internal and external stability. Those perturbations are small by their norms⁶.

In order to formalize the condition of stable realizability it is convenient to use the drift equations in the form (14.66). Let the distribution μ^* be internally and externally stable in the first approximations. Let the points $x^{*\alpha}$ of global maxima of $k_{\mu^*}(x)$ be non-degenerate in the second approximation. This means that the matrices

$$b_{ij}^{*\alpha} = - \left(\frac{\partial^2 k_{\mu^*}(x)}{\partial x_i \partial x_j} \right)_{x=x^{*\alpha}} \quad (14.73)$$

are strictly positively definite for all α .

Under these conditions of stability and non-degeneracy the coefficients of (14.66) can be easily calculated using Taylor series expansion in powers of $(x^\alpha - x^{*\alpha})$. The stable realizability of μ^* in the first approximation means that the fixed point of the drift equations (14.66) with the coordinates

$$x^\alpha = x^{*\alpha}, \quad b_{ij}^\alpha = b_{ij}^{*\alpha} \quad (14.74)$$

is stable in the first approximation. It is the usual stability for the system (14.66) of ordinary differential equations.

14.4.8 Main Results About Systems with Inheritance

1. If a kinetic equation has the quasi-biological form (14.45) then it has a rich system of invariant manifolds: for any closed subset $A \subset X$ the set of distributions $\mathbf{M}_A = \{\mu \mid \text{supp} \mu \subseteq A\}$ is invariant with respect to the system (14.45). These invariant manifolds form important algebraic structure, the summation of manifolds is possible:

$$\mathbf{M}_A \oplus \mathbf{M}_B = \mathbf{M}_{A \cup B}.$$

(Of course, $\mathbf{M}_{A \cap B} = \mathbf{M}_A \cap \mathbf{M}_B$).

⁶ Let us remind that the norm of the measure μ is $\|\mu\| = \sup_{|f| \leq 1} \mu[f]$. If one shifts the δ -measure of unite mass by any nonzero distance ε , then the norm of the perturbation is 2. Nevertheless, this perturbation weakly tends to 0 with $\varepsilon \rightarrow 0$.

2. Typically, all the ω -limit points belong to invariant manifolds \mathcal{M}_A with finite A . The finite-dimensional approximations of the reproduction coefficient (14.56) provides the minimax estimation of the number of points in A .
3. For systems with inheritance (14.45) a solution typically tends to be a finite set of narrow peaks getting in time more and more narrow, moving slower and slower. It is possible that these peaks do not tend to fixed positions, rather they continue moving, and the path covered tends to infinity at $t \rightarrow \infty$. This is the *drift effect*.
4. The equations for peak dynamics, the drift equations, (14.65,14.66,14.69) describe dynamics of the shapes of the peaks and their positions. For systems with small variability (“mutations”) the drift equations (14.69) has the scaling property: the change of the intensity of mutations is equivalent to the change of the time scale.
5. Three specific types of stability are important for the systems with inheritance: internal stability (stability with respect to perturbations without extension of distribution support), external stability (stability with respect to small one-point extension of distribution support), and stable realizability (stability with respect to weakly small⁷ perturbations: small extensions and small shifts of the peaks).

Some exact results of the mathematical selection theory can be found in [402,403]. There exist many physical examples of systems with inheritance [385–391]. A wide field of ecological applications was described in the book [383]. An introduction into adaptive dynamics was given in notes [404] that illustrate largely by way of examples, how standard ecological models can be put into an evolutionary perspective in order to gain insight in the role of natural selection in shaping life history characteristics. The cell division self-synchronization below demonstrates effects of unusual inherited unit, it is the example of a “phase selection”.

14.5 Example: Cell Division Self-Synchronization

The results described above admit for a whole family of generalizations. In particular, it seems to be important to extend the theorems of selection to the case of vector distributions, when $k_\mu(x)$ is a linear operator at each μ, x . It is possible also to make generalizations for some classes of non-autonomous equations with explicit dependencies of $k_\mu(x)$ on t .

Availability of such a network of generalizations allows to construct the reasoning as follows: *what* is inherited (i.e. for what the law of conservation of support holds) is the subject of selection (i.e. with respect to these variables at $t \rightarrow \infty$ the distribution becomes discrete and the limit support can be described by the optimality principles).

⁷ That is, small in the weak topology.

This section gives a somewhat unconventional example of inheritance and selection, when the reproduction coefficients are subject additional conditions of symmetry.

Consider a culture of microorganisms in a certain medium (for example, pathogenous microbes in the organism of a host). Assume, for simplicity, the following: *let the time period spent by these microorganisms for the whole life cycle be identical.*

At the end of the life cycle the microorganism disappears and new several microorganisms appear in the initial phase. Let T be the time of the life cycle. Each microorganism holds the value of the inherited variable, it is “the moment of its appearance (mod T)”. Indeed, if the given microorganism emerges at time τ ($0 < \tau \leq T$), then its first descendants appear at time $T + \tau$, the next generation – at the moment $2T + \tau$, then $3T + \tau$ and so on.

It is natural to assume that the phase $\tau \pmod T$ is the inherited variable. This implies selection of phases and, therefore, survival of their discrete number τ_1, \dots, τ_m , only. But results of the preceding sections cannot be applied directly to this problem. The reason is the additional symmetry of the system with respect to the phase shift. But the typicalness of selection and the instability of the uniform distribution over the phases $\tau \pmod T$ can be shown for this case, too. Let us illustrate it with the simplest model.

Let the difference between the microorganisms at each time moment be related to the difference in the development phases only. Let us also assume that the state of the medium can be considered as a function of the distribution $\mu(\tau)$ of microorganisms over the phases $\tau \in]0, T]$ (the quasi-steady state approximation for the medium). Consider the system at discrete times nT and assume the coefficient connecting μ at moments nT and $nT + T$ to be the exponent of the linear integral operator value:

$$\mu_{n+1}(\tau) = \mu_n(\tau) \exp \left[k_0 - \int_0^T k_1(\tau - \tau') \mu_n(\tau') d\tau' \right]. \quad (14.75)$$

Here, $\mu_n(\tau)$ is the distribution at the moment nT , $k_0 = \text{const}$, $k_1(\tau)$ is a periodic function of period T .

The uniform steady-state $\mu^* \equiv n^* = \text{const}$ is:

$$n^* = \frac{k_0}{\int_0^T k_1(\theta) d\theta}. \quad (14.76)$$

In order to examine stability of the uniform steady state μ^* (14.76), the system (14.75) is linearized. For small deviations $\Delta\mu(\tau)$ in linear approximation

$$\Delta\mu_{n+1}(\tau) = \Delta\mu_n(\tau) - n^* \int_0^T k_1(\tau - \tau') \Delta\mu_n(\tau') d\tau'. \quad (14.77)$$

Expand $k_1(\theta)$ into the Fourier series:

$$k_1(\theta) = b_0 + \sum_{n=1}^{\infty} \left(a_n \sin \left(2\pi n \frac{\theta}{T} \right) + b_n \cos \left(2\pi n \frac{\theta}{T} \right) \right). \quad (14.78)$$

Denote by A operator of the right-hand side of (14.77). In the basis of functions

$$e_{s_n} = \sin \left(2\pi n \frac{\theta}{T} \right), \quad e_{c_n} = \cos \left(2\pi n \frac{\theta}{T} \right)$$

on the segment $]0, T]$ the operator A is block-diagonal. The vector e_0 is eigenvector, $Ae_0 = \lambda_0 e_0$, $\lambda_0 = 1 - n^* b_0 T$. On the two-dimensional space, generated by vectors e_{s_n}, e_{c_n} the operator A is acting as a matrix

$$A_n = \begin{pmatrix} 1 - \frac{Tn^*}{2} b_n & -\frac{Tn^*}{2} a_n \\ \frac{Tn^*}{2} a_n & 1 - \frac{Tn^*}{2} b_n \end{pmatrix}. \quad (14.79)$$

The corresponding eigenvalues are

$$\lambda_{n,1,2} = 1 - \frac{Tn^*}{2} (b_n \pm ia_n). \quad (14.80)$$

For the uniform steady state μ^* (14.76) to be unstable it is sufficient that the absolute value of at least one eigenvalue $\lambda_{n,1,2}$ be larger than 1: $|\lambda_{n,1,2}| > 1$. If there is at least one negative Fourier cosine-coefficient $b_n < 0$, then $\text{Re}\lambda_n > 1$, and thus $|\lambda_n| > 1$.

Note now, that almost all periodic functions (continuous, smooth, analytical – this does not matter) have negative Fourier cosine-coefficient. This can be understood as follows. The sequence b_n tends to zero at $n \rightarrow \infty$. Therefore, if all $b_n \geq 0$, then, by changing b_n at sufficiently large n , we can make b_n negative, and the perturbation value can be chosen less than any previously set positive number. On the other hand, if some $b_n < 0$, then this coefficient cannot be made non-negative by sufficiently small perturbations. Moreover, the set of functions that have all Fourier cosine-coefficient non-negative is completely thin, because for any compact of functions K (for most of norms in use) the sequence $B_n = \max_{f \in K} |b_n(f)|$ tends to zero, where $b_n(f)$ is the n th Fourier cosine-coefficient of function f .

The model (14.75) is revealing, because for it we can trace the dynamics over large times, if we restrict ourselves with a finite segment of the Fourier series for $k_1(\theta)$. Describe it for

$$k_1(\theta) = b_0 + a \sin \left(2\pi \frac{\theta}{T} \right) + b \cos \left(2\pi \frac{\theta}{T} \right). \quad (14.81)$$

Assume further that $b < 0$ (then the homogeneous distribution $\mu^* \equiv \frac{k_0}{b_0 T}$ is unstable) and $b_0 > \sqrt{a^2 + b^2}$ (then the $\int \mu(\tau) d\tau$ cannot grow unbounded in time). Introduce notations

$$\begin{aligned}
M_0(\mu) &= \int_0^T \mu(\tau) d\tau, \quad M_c(\mu) = \int_0^T \cos\left(2\pi\frac{\tau}{T}\right) \mu(\tau) d\tau, \\
M_s(\mu) &= \int_0^T \sin\left(2\pi\frac{\tau}{T}\right) \mu(\tau) d\tau, \quad \langle\mu\rangle_n = \frac{1}{n} \sum_{m=0}^{n-1} \mu_m, \quad (14.82)
\end{aligned}$$

where μ_m is the distribution μ at the discrete time m .

In these notations,

$$\begin{aligned}
\mu_{n+1}(\tau) &= \mu_n(\tau) \exp \left[k_0 - b_0 M_0(\mu_n) - (aM_c(\mu_n) + bM_s(\mu_n)) \sin\left(2\pi\frac{\tau}{T}\right) \right. \\
&\quad \left. + (aM_s(\mu_n) - bM_c(\mu_n)) \cos\left(2\pi\frac{\tau}{T}\right) \right]. \quad (14.83)
\end{aligned}$$

Represent the distribution $\mu_n(\tau)$ through the initial distribution $\mu_0(\tau)$ and the functionals M_0, M_c, M_s values for the average distribution $\langle\mu\rangle_n$:

$$\begin{aligned}
\mu_n(\tau) &= \mu_0(\tau) \\
&\times \exp \left\{ n \left[k_0 - b_0 M_0(\langle\mu\rangle_n) - (aM_c(\langle\mu\rangle_n) + bM_s(\langle\mu\rangle_n)) \sin\left(2\pi\frac{\tau}{T}\right) \right. \right. \\
&\quad \left. \left. + (aM_s(\langle\mu\rangle_n) - bM_c(\langle\mu\rangle_n)) \cos\left(2\pi\frac{\tau}{T}\right) \right] \right\}. \quad (14.84)
\end{aligned}$$

The exponent in (14.84) is either independent of τ , or there is a function with the single maximum on $]0, T[$. The coordinate $\tau_n^\#$ of this maximum is easily calculated

$$\tau_n^\# = -\frac{T}{2\pi} \arctan \frac{aM_c(\langle\mu\rangle_n) + bM_s(\langle\mu\rangle_n)}{aM_s(\langle\mu\rangle_n) - bM_c(\langle\mu\rangle_n)} \quad (14.85)$$

Let the non-uniform smooth initial distribution μ_0 has the whole segment $[0, T]$ as its support. At the time progress the distributions $\mu_n(\tau)$ takes the shape of ever narrowing peak. With high accuracy at large a we can approximate $\mu_n(\tau)$ by the Gaussian distribution (approximation accuracy is understood in the weak sense, as closeness of mean values):

$$\begin{aligned}
\mu_n(\tau) &\approx M_0 \sqrt{\frac{q_n}{\pi}} \exp[-q_n(\tau - \tau_n^\#)^2], \quad M_0 = \frac{k_0}{k_1(0)} = \frac{k_0}{b_0 + b}, \quad (14.86) \\
q_n^2 &= n^2 \left(\frac{2\pi}{T}\right)^4 \left[(aM_c(\langle\mu\rangle_n) + bM_s(\langle\mu\rangle_n))^2 + (aM_s(\langle\mu\rangle_n) - bM_c(\langle\mu\rangle_n))^2 \right].
\end{aligned}$$

Expression (14.86) involves the average measure $\langle\mu\rangle_n$ which is difficult to compute. However, we can operate without direct computation of $\langle\mu\rangle_n$. At $q_n \gg \frac{1}{T^2}$ we can compute q_{n+1} and $\tau_{n+1}^\#$:

$$\begin{aligned}
\mu_{n+1} &\approx M_0 \sqrt{\frac{q_n + \Delta q}{\pi}} \exp[-(q_n + \Delta q)((\tau - \tau_n^\# - \Delta\tau^\#)^2)], \\
\Delta q &\approx -\frac{1}{2} b M_0 \left(\frac{2\pi}{T}\right)^2, \quad \Delta\tau^\# \approx \frac{1}{q} M_0 \frac{2\pi}{T}. \quad (14.87)
\end{aligned}$$

The accuracy of these expression grows with time n . The value q_n grows at large n almost linearly, and $\tau_n^\#$, respectively, as the sum of the harmonic series (mod T), i.e. as $\ln n \pmod T$. The drift effect takes place: location of the peak $\tau_n^\#$, passes at $n \rightarrow \infty$ the distance diverging as $\ln n$.

Of interest is the case, when $b > 0$ but

$$|\lambda_1|^2 = \left(1 + n^*b\frac{T}{2}\right)^2 + \left(n^*a\frac{T}{2}\right)^2 > 1.$$

With this, homogeneous distribution $\mu^* \equiv n^*$ is not stable but μ does not tend to δ -functions. There are smooth stable “self-synchronization waves” of the form

$$\mu_n = \gamma \exp \left[q \cos \left((\tau - n\Delta\tau^\#) \frac{2\pi}{T} \right) \right].$$

At small $b > 0$ ($b \ll |a|, bM_0 \ll a^2$) we can find explicit form of approximated expressions for q and $\Delta\tau^\#$:

$$q \approx \frac{a^2 M_0}{2b}, \quad \Delta\tau^\# \approx \frac{bT}{\pi a}. \tag{14.88}$$

At $b > 0, b \rightarrow 0$, smooth self-synchronization waves become ever narrowing peaks, and their steady velocity approaches zero. If $b = 0, |\lambda_1|^2 > 1$ then the effect of selection takes place again, and for almost all initial conditions μ_0 with the support being the whole segment $[0, T]$ the distribution μ_n takes

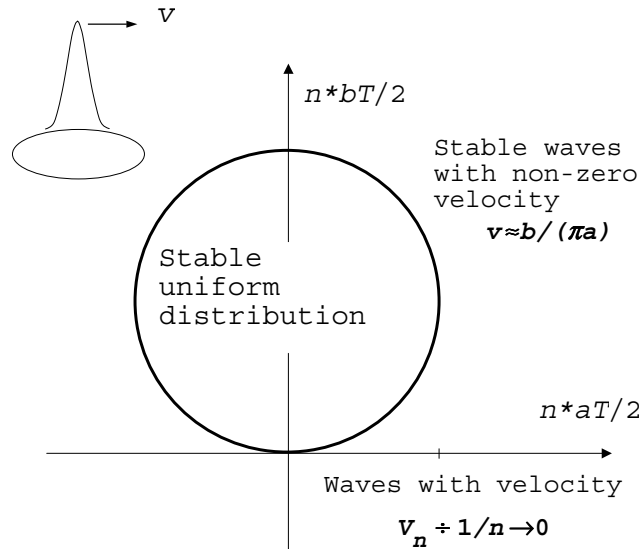


Fig. 14.1. The simplest model of cell division self-synchronization: The parametric portrait

at large n the form of a slowly drifting almost Gaussian peak. It becomes narrower with the time, and the motion slows down. Instead of the linear growth of q_n which takes place at $b < 0$ (14.87), for $b = 0$, $q_{n+1} - q_n \approx \text{const}q_n^{-1}$ and q_n grows as $\text{const}\sqrt{n}$.

The parametric portrait of the system for the simple reproduction coefficient (14.81) is presented in Fig. 14.1.

As usual, a small desynchronization transforms δ -functional limit peaks to narrow Gaussian peaks, and the velocity of peaks tends to small but nonzero velocity instead of zero. The systems with small desynchronization can be described by equations of the form (14.69).

There are many specific mechanisms of synchronization and desynchronization in physics and biology (see, for example [405–409]). We described here very simple mechanism: it requires only that the time of the life cycle is fixed, in this case in a generic situation we should observe the self-synchronization. Of course, the real-world situation can be much more complicated, with a plenty of additional factors, but the basic mechanism of the “phase selection” works always if the life cycle has more or less fixed duration.