

Research



Cite this article: Shalymov DS, Fradkov AL.

2016 Dynamics of non-stationary processes that follow the maximum of the Rényi entropy principle. *Proc. R. Soc. A* **472**: 20150324. <http://dx.doi.org/10.1098/rspa.2015.0324>

Received: 16 May 2015

Accepted: 22 December 2015

Subject Areas:

mathematical physics, cybernetics, computer modelling and simulation

Keywords:

Rényi entropy, maximum entropy principle, Rényi distribution, speed-gradient principle

Author for correspondence:

Dmitry S. Shalymov

e-mail: dmitry.shalymov@gmail.com

Dynamics of non-stationary processes that follow the maximum of the Rényi entropy principle

Dmitry S. Shalymov¹ and Alexander L. Fradkov^{1,2}

¹Mathematics and Mechanics Faculty, St Petersburg State University, St Petersburg, Russia

²The Laboratory 'Control of Complex Systems', Institute of Problems of Mechanical Engineering RAS, St Petersburg, Russia

DSS, 0000-0002-8794-7306; ALF, 0000-0002-5633-0944

We propose dynamics equations which describe the behaviour of non-stationary processes that follow the maximum Rényi entropy principle. The equations are derived on the basis of the speed-gradient principle originated in the control theory. The maximum of the Rényi entropy principle is analysed for discrete and continuous cases, and both a discrete random variable and probability density function (PDF) are used. We consider mass conservation and energy conservation constraints and demonstrate the uniqueness of the limit distribution and asymptotic convergence of the PDF for both cases. The coincidence of the limit distribution of the proposed equations with the Rényi distribution is examined.

1. Introduction

Entropy is actively used by many fields of science including physics, chemistry, computer science, biology, etc. [1]

Today there are many different types of entropy in use, which often become the centre of discussions in both statistical physics and thermodynamics. The most famous is the Shannon entropy [2]:

$$H(X) = - \sum_i P(x_i) \log P(x_i), \quad (1.1)$$

where X is a discrete random variable with possible values $\{x_1, \dots, x_n\}$ and P is a probability mass function.

In 1961, Rényi [3] introduced a generalized Shannon entropy as a one-parameter family of entropy

$$H(X, \beta) = \frac{1}{1 - \beta} \log \left(\sum_i p^\beta(x_i) \right), \quad \beta \neq 1, \beta > 0. \quad (1.2)$$

The Rényi entropy tends to the Shannon entropy when $\beta \rightarrow 1$.

Extension of the Rényi entropy for the continuous case can be defined as

$$H(X, \beta) = \frac{1}{1 - \beta} \log \left(\int p^\beta(x) dx \right), \quad \beta \neq 1, \beta > 0, \quad (1.3)$$

where X is an absolutely continuous random variable with probability density function (PDF) p .

The Rényi entropy is often taken as a typical measure of complexity to describe dynamical systems in physics, engineering and information theory [4]. The Rényi entropy applications are overviewed in [5].

A variety of physical systems obey the famous maximum entropy (MaxEnt) principle: their entropy achieves maximum under constraints caused by other physical laws. Since 1957, when the seminal works of Jaynes were published [6–8], and until now [9–11], the MaxEnt principle has caused lively interest among researchers. For example, the MaxEnt for the Rényi entropy has been successfully applied to generalize the Thomas–Fermi model in [12].

Although the states of maximum entropy are widely discussed in scientific articles and studies, the dynamics of evolution and transient behaviour of systems are still not well investigated.

In this paper, we propose a set of equations that describe the dynamics of the PDFs for non-stationary processes that follow the maximum of the Rényi entropy principle.

The speed-gradient (SG) principle [13–16] used here is originated from control theory. This principle has already been applied in [13,15,17] to derive the equations of dynamics for systems with a finite number of particles for the case of the maximum Shannon entropy principle. Systems with continuous probability distributions are considered in [18]. The dynamics of discrete systems for the Tsallis entropy is discussed in [19]. The SG principle for systems with a finite number of particles is studied in [15], where systems are emulated with the molecular dynamics method. We take a similar approach for systems with discrete and continuous probability distributions while researching the Rényi entropy. The derived equations describe the dynamics of non-stationary (transient) states and show the way and trajectory of a system tending to the state with maximum Rényi entropy.

The well-known Fokker–Planck (FP) equations [20] describe the time evolution of the PDF. Jaynes’s MaxEnt approach can also be applied to these equations [21,22]. Another general form of time-evolution equations for non-equilibrium systems is known as GENERIC (general equation for the non-equilibrium reversible–irreversible coupling) [23,24]. The relation between GENERIC and FP equations is established in [23]. It states that FP equations are a particular case of the GENERIC when a noise term is added into the GENERIC. Thus, if the FP equation is represented as a stochastic differential equation, from which fluctuations are eliminated, this equation matches the GENERIC equation [25].

Following this, we can also claim that the SG principle matches the GENERIC (and thus it matches the FP equation) if a goal function is set as entropy and constraints are specified by energy [19]. Moreover, the SG principle is a more general case of the GENERIC because almost every smooth function can be taken as a goal function, not only entropy. The GENERIC and the SG principle relations are examined in §3.

We propose an evolution law of the system in the following form:

$$\dot{p} = -\Gamma(I - \Psi)p^{\beta-1}, \quad (1.4)$$

where I is an identity operator, Ψ is a linear integral operator that does not depend on p and $\Gamma > 0$ is a constant gain. All the solutions of equation (1.4) converge to the distribution that corresponds to the maximum value of the Rényi entropy.

A way to find the distribution achieving the extreme value (maximum or minimum) of entropy within a given variational distance from any given distribution is proposed in [26]. An approximation of the probability distribution is built there based on new bounds of entropy distance. These bounds are also applied for the entropy estimation in [27].

Jaynes's MaxEnt principle was successfully used for inductive inference in [28]. It is shown that, given information in the form of constraints on expected values, there is only one appropriate distribution which can be obtained by maximizing entropy. Inductive inference can be used in systems with non-stationary processes which satisfy the MaxEnt [29]. In [29], the notion of entropy dynamics (ED) is used. ED is a theoretical framework which combines inductive inference [28] and information geometry [30]. ED investigates the possibility of deriving dynamics from purely entropic arguments.

Besides the Rényi entropy, more general forms of relative entropies and divergences can also be studied from a perspective of the SG principle, for example the Cressie–Read and the Csiszár–Morimoto conditional entropies (f-divergences) [31].

This paper has the following structure. The next section describes the SG principle, accompanied with two examples. The third section examines the relations between the SG principle, the GENERIC and the FP equations. The fourth section introduces Jaynes's formalism; the Rényi distribution (RD) is derived from the maximum of the Rényi entropy principle. The fifth section gives an example of a dynamic system with discrete distribution of parameters that follows the maximum of the Rényi entropy principle (we consider cases with one and two constraints and derive equations for the transient states). The sixth section extends the results obtained in the fifth section for the case of the continuous PDF. The asymptotic stability of PDF dynamics is proved for the cases with one and two constraints.

2. The speed-gradient principle

There is a connection between the laws of control in technical systems and the laws of dynamics in physical systems. It is known that the methods for the synthesis of control algorithms allow one to derive the laws of dynamics for physical systems. In particular, the model of the dynamics for a number of physical systems can be derived based on the SG method with an appropriate choice of goal function.

Consider the class of open physical systems whose dynamics can be described by the system of differential equations

$$\dot{x} = f(x, u, t), \quad (2.1)$$

where $x \in \mathbb{C}^n$ is the system state vector and u is the vector of input (free) variables $t \geq 0$. The problem of modelling of the system can be formulated as finding the law of change (evolution) $u(t)$ which meets a certain criterion of 'naturalness' for its behaviour and grants a set of properties observed in real physical systems to a generated model.

Such formulations are well known in physics. Variational principles of systems models have long been recognized. They usually involve the task of an integral functional that characterizes the behaviour of the system [32]. Minimization of the functional defines the real possible trajectories of the system $\{x(t), u(t)\}$ as points in the corresponding functional space. To explicitly specify the dynamics of the system a developed apparatus of the variations calculus is used.

Methods of optimal control (e.g. Bellman dynamic programming, Pontryagin maximum principle, etc.) are the result of the development of classical variational calculus methods. And it can be used to build dynamic models of mechanical systems in Nature and society.

Together with integral principles, differential local time principles have also been proposed, such as the Gauss principle of least constraint, the principle of minimum energy dissipation, etc. As noted by Planck [33], local principles have some advantages over integral ones, because they do not make the current state and the movement of the system dependent on later states and movements. Let us formulate another local variational principle based on the method of SG [14, 17], as follows.

The speed-gradient principle. *Only those possible movements of the system are realized (among all possible movements) for which the input variables change proportionally to the SG of a 'goal' functional.*

The SG principle offers researchers the choice of two types of systems dynamics models:

(A) models which follow the algorithm of the SG in differential form:

$$\dot{u} = -\Gamma \nabla_u \dot{Q}_t; \quad (2.2)$$

(B) models following the algorithm of the SG in finite form:

$$u = -\Gamma \nabla_u \dot{Q}_t, \quad (2.3)$$

where \dot{Q}_t is the rate of change of the target functional along the trajectories of the system (2.1). We describe the application of the SG principle in the simplest (and most important) case when the class of models of dynamics is given by the relation:

$$\dot{x} = u. \quad (2.4)$$

Relation (2.4) means only that we are looking for a law of variation rates of the state variables of the system.

In accordance with the SG principle, the goal functional $Q(x)$ has to be determined first. Selection of $Q(x)$ should be based on the physics of the real system and reflect the presence of a tendency to decrease the current value of $Q(x(t))$. After that, the law of dynamics can be written in the form (2.2) or (2.3).

Let us illustrate the introduced SG principle with several examples.

(a) Example 1: the motion of a particle in a potential field

As a first example, consider the problem of describing the motion of a particle in a potential field. State variables here are the coordinates of the point x_1, x_2, x_3 which form a vector $x = (x_1, x_2, x_3)^T$.

We choose a smooth 'goal' functional $Q(x)$ as the potential energy of a particle and derive the SG law in the differential form. We calculate the speed

$$\dot{Q} = \frac{\partial}{\partial t} Q = [\nabla_x Q(x)]^T u \quad (2.5)$$

and the SG

$$\nabla_u \dot{Q} = \nabla_x Q(x). \quad (2.6)$$

Then, choosing the diagonal positive definite gain matrix $\Gamma = m^{-1} I_3$, where $m > 0$ is a parameter and I_3 is the 3×3 identity matrix, we arrive at Newton's law $\dot{u} = -m^{-1} \nabla_x Q(x)$ or

$$m\ddot{x} = -\nabla_x Q(x),$$

where m is interpreted as the mass of a particle.

Note that the SG laws with non-diagonal gain matrices Γ can be incorporated if a non-Euclidean metric in the space of inputs is introduced by the matrix Γ^{-1} . Admitting dependence of the metric matrix Γ on x , one can obtain evolution laws for complex mechanical systems described by Lagrangian or Hamiltonian formalism.

The SG principle applies not only to finite-dimensional systems but also to infinite-dimensional (distributed) ones. Particularly, x may be a vector of a functional space \mathcal{X} and $f(x, u, t)$ may be a nonlinear differential operator (in such a case, the solutions of (2.1) should be understood as generalized ones). We will omit the mathematical details for simplicity.

(b) Example 2: the viscous fluid dynamics

Let the infinite-dimensional state vector be formed of two functions: $x = (v(\cdot, t), p(\cdot, t))^T$, where $v(r, t) \in \mathbb{R}^3$ is the velocity field of the three-dimensional fluid flow and $p(r, t)$ is the pressure field.

We introduce the ‘goal’ functional in the following way:

$$Q_t = \int_{\Omega} p(r, t) dr + v_0 \int_{\Omega} |\nabla_r v(r, t)|^2 dr, \quad (2.7)$$

where $v_0 > 0$ is a weight coefficient. Calculating functional (2.7) relative to (2.4) gives $\nabla_u \dot{Q}_t = \nabla_r p - v_0 \Delta v$. The differential form of the SG principle is the Navier–Stokes equation, which describes the motion of a viscous fluid:

$$\rho \frac{\partial v}{\partial t}(r, t) = -\nabla_r p(r, t) + v_0 \Delta v(r, t),$$

where $v = v_0 \gamma^{-1}$ is the viscosity factor and $\rho = \gamma^{-1}$ is the density of the liquid.

Note that the differential form of the SG laws often corresponds to reversible processes, while the finite form generates irreversible ones. For modelling of more complex dynamics, a combination of finite and differential SG laws may be useful.

In a similar way, dynamical equations for many other mechanical, electrical and thermodynamic systems can be recovered. The SG principle applies to a broad class of physical systems subjected to potential and/or dissipative forces.

This paper is aimed at application of the SG principle to entropy-driven systems.

3. GENERIC and the speed-gradient principle

The GENERIC time-evolution equation is formulated as [23]:

$$\frac{dx}{dt} = L(x) \frac{\partial E(x)}{\partial x} + M(x) \frac{\partial S(x)}{\partial x}, \quad (3.1)$$

where x is a state of a system, E is a total energy functional and S is an entropy functional.

Let us show the relation between the GENERIC and SG equations. Assume that we have to maximize the entropy function $S(x)$ of a system that has an additional constraint for a total energy $E(x) = E = \text{const}$. The Lagrangian for this case can be defined as:

$$\Lambda(x, \lambda) = S(x) + \lambda(E(x) - E), \quad (3.2)$$

where λ is a real Lagrange multiplier.

Let us use the SG principle equation (2.3) for the Lagrangian (3.2), i.e. $Q_t = \Lambda$,

$$u = -\Gamma \nabla_u \dot{\Lambda}(x, \lambda) = -\Gamma \nabla_u [\dot{S} + \lambda \dot{E}]. \quad (3.3)$$

According to (2.5) and (2.6) we can write equation (3.3) as:

$$u = -\Gamma [\nabla_x S(x) + \lambda \nabla_x E(x)]. \quad (3.4)$$

We can see that the dynamics equation obtained from the SG principle (3.4) coincides with the GENERIC equation (3.1) for $dx/dt = u$, $L(x) = -\Gamma \lambda$ and $M(x) = -\Gamma$.

GENERIC is based on two ‘potentials’ of total energy and entropy. The SG principle can use any smooth functional that has to be maximized (minimized) as a goal function. So it is not necessary to be only Lagrangian (3.2) or entropy functional. The SG principle can be treated as a more general approach to describe the dynamics of a system. Nevertheless, GENERIC is also a general equation. It uses parametrized matrices $L(x)$ and $M(x)$ that make it possible to use GENERIC for a wide range of time-evolution systems.

The relation between GENERIC and FP equations is established in [23], which means that there is also a relation between the FP equations and the SG principle. So many different methods allow the researcher to select the one that is most convenient or more fully describes the problem being solved.

4. Maximum entropy principle

(a) Jaynes's MaxEnt

Jaynes [6–8] proposed the approach which became one of the foundations for modern-day statistical physics.

Let $p(x)$ be an unknown PDF of a multi-dimensional random variable x . Suppose that we have to define it on the basis of certain system information. Consider a continuous case and suppose that there is *a priori* known information about some average values \bar{H}_m ,

$$\bar{H}_m = \int H_m(x)p(x) dx, \quad m = 1, \dots, M. \quad (4.1)$$

Conditions (4.1) can be insufficient to derive $p(x)$ in general. According to Jaynes, applying maximization of information entropy $H(X)$ is the most objective method to define the PDF in this case.

Lagrange multipliers are used to perform the maximum entropy search with additional conditions (4.1). This leads to

$$p(x) = \frac{1}{Z} \exp \left(- \sum_{m=1}^M \lambda_m H_m(x) \right), \quad (4.2)$$

where $Z = \int \exp(-\sum_{m=1}^M \lambda_m H_m(x)) dx$ and λ_m can be derived from conditions (4.1).

These formulae show the match between the maximum information entropy and the Gibbs entropy in the case of equilibrium. So the information entropy can be identified with the thermodynamic entropy in this case.

As the Rényi entropy is a generalization of the Shannon entropy, it seems natural to extend Jaynes's maximum entropy principle to the case of the Rényi entropy.

(b) The Rényi distribution

The RD that corresponds to the state with maximum Rényi entropy is proposed in [34,35]. The RD is considered there for a system with a discrete probability distribution when two constraints are imposed: normalization constraint (mass conservation law) $\sum_i p_i = 1$ and a fixed average energy (total energy) constraint $\sum_i H_i p_i = E$.

If the Rényi entropy is used instead of the Shannon entropy in MaxEnt, then the equilibrium distribution (the RD) can be formulated as

$$p_i = \frac{1}{Z_R} \left(1 - \lambda \frac{\beta - 1}{\beta} (H_i - E) \right)^{1/(\beta-1)}, \quad (4.3)$$

where $Z_R = \sum_i (1 - \lambda((\beta - 1)/\beta)(H_i - E))^{1/(\beta-1)}$ and λ is a Lagrange multiplier which can be derived from the total energy constraint.

Following the constraints in form (4.1), the expression (4.3) for continuous distributions becomes

$$p(x) = \frac{1}{Z_R} \left(1 - \lambda \frac{\beta - 1}{\beta} (h(x) - E) \right)^{1/(\beta-1)}, \quad (4.4)$$

where $Z_R = \int (1 - \lambda((\beta - 1)/\beta)(h(x) - E))^{1/(\beta-1)} dx$ and λ is a Lagrange multiplier.

5. The speed-gradient dynamics of the Rényi entropy maximization process

We extend the approach introduced in [13] to the case of the Rényi entropy. Similar to [13] for the Shannon entropy, we consider a discrete system which consists of N identical particles distributed over m cells.

In the case when the mass conservation constraint holds, it is true that $\sum_{i=1}^m N_i = N$. It can be normalized as

$$\sum_{i=1}^m \frac{N_i}{N} = 1. \quad (5.1)$$

Particles can move from one cell to another. The steady-state and the transient behaviour of the system are both interesting for us. According to the MaxEnt principle, the limit behaviour of the system maximizes its entropy for the steady state when nothing else is known [7,8].

To obtain a transient mode behaviour, we apply the SG principle, choosing the Rényi entropy as the goal function to be maximized

$$H(X, \beta) = \frac{1}{1 - \beta} \log \left(\sum_{i=1}^m N_i^\beta \right), \quad (5.2)$$

where $\beta \geq 0$, $\beta \neq 1$ and $X = (N_1, \dots, N_m)^T$ is the state vector of the system.

Assume that the motion is continuous in time and the numbers N_i are changing continuously. Then the law of motion can be represented as $\dot{N}_i = u_i$, $i = 1, \dots, m$, where $u_i = u_i(t)$ are control functions which have to be determined.

The evaluation scheme is as follows. First the speed of entropy change is evaluated:

$$\dot{H}(\beta) = \frac{\beta}{1 - \beta} \frac{\sum_{i=1}^m N_i^{\beta-1} \dot{N}_i}{\sum_{i=1}^m N_i^\beta}.$$

Then the gradient of the speed is evaluated with respect to the vector of controls u_i considered as frozen parameters,

$$\nabla_{u_i} \dot{H}(\beta) = \frac{\beta}{1 - \beta} \frac{N_i^{\beta-1}}{\sum_{i=1}^m N_i^\beta}.$$

And finally the actual controls are defined proportional to the projection of the SG to the surface of constraints (5.1),

$$\dot{N}_i = -\Gamma \frac{\beta}{1 - \beta} \frac{N_i^{\beta-1}}{\sum_{i=1}^m N_i^\beta} + \lambda',$$

where λ' is a Lagrange multiplier chosen in order to fulfil the constraint (5.1).

Now we can evaluate λ' :

$$\sum_{i=1}^m \dot{N}_i = 0 \Rightarrow \lambda' = \frac{\Gamma}{m} \frac{\beta}{1 - \beta} \frac{\sum_{i=1}^m N_i^{\beta-1}}{\sum_{i=1}^m N_i^\beta}.$$

The final form of the system dynamics law is as follows:

$$\dot{N}_i = -\Gamma \frac{\beta}{1 - \beta} \frac{N_i^{\beta-1}}{\sum_{i=1}^m N_i^\beta} + \frac{\Gamma}{m} \frac{\beta}{1 - \beta} \frac{\sum_{i=1}^m N_i^{\beta-1}}{\sum_{i=1}^m N_i^\beta}. \quad (5.3)$$

Let us find the equilibrium mode which corresponds to the asymptotic behaviour of the variables N_i . In this mode $\dot{N}_i = 0$. Based on (5.3), this means that $m N_i^{\beta-1} = \sum_{i=1}^m N_i^{\beta-1}$. Hence all N_i are equal. According to constraint (5.1), we have that $N_i = N/m$. This result corresponds to RD (4.3) for the case of one constraint (i.e. $\lambda = 0$ as there is no total energy constraint). It also corresponds to the maximum state of classical entropy and agrees with thermodynamics.

(a) Equilibrium stability

Let us examine the stability of the equilibrium mode. We introduce the Lyapunov function

$$V(X, \beta) = H_{\max}(\beta) - H(X, \beta).$$

Evaluation of \dot{V} yields

$$\begin{aligned} \dot{V}(\beta) &= -\dot{H}(X, \beta) = \frac{\beta}{1 - \beta} \frac{\sum_{i=1}^m N_i^{\beta-1} u_i}{\sum_{i=1}^m N_i^\beta} \\ &= \Gamma \left(\frac{\beta}{1 - \beta} \right)^2 \frac{1}{(\sum_{i=1}^m N_i^\beta)^2} \left(\left(\sum_{i=1}^m N_i^{\beta-1} \right)^2 - m \sum_{i=1}^m (N_i^{\beta-1})^2 \right). \end{aligned}$$

Consider the Cauchy–Bunyakovsky–Schwarz (CBS) inequality for two vectors $a, b \in \mathbb{R}^m$:

$$\left(\sum_{i=1}^m a_i b_i \right)^2 \leq \sum_{i=1}^m a_i^2 \sum_{i=1}^m b_i^2. \quad (5.4)$$

According to the CBS inequality for vectors $a = (N_1^{\beta-1}, \dots, N_m^{\beta-1})$ and $b = (1, \dots, 1)$, we have that $\dot{V}(\beta) \leq 0$. Equality $\dot{V}(\beta) = 0$ holds if and only if all values N_i are equal. This is the maximum entropy state. Thus law (5.3) provides global asymptotic stability of the maximum entropy state. The physical meaning of this law is nothing but moving along the direction of the maximum entropy production rate (direction of the fastest entropy growth).

(b) Total energy constraint

The case of more than one constraint can be treated in the same way. Suppose that in addition to the mass conservation law (5.1) the energy conservation law also holds. Let E_i be the energy of the particle in the i th cell and the total energy does not change. The total energy constraint is

$$E = \sum_{i=1}^m \frac{N_i}{N} E_i. \quad (5.5)$$

A new set of constraints can be formulated as

$$\sum_{i=1}^m u_i = 0 \quad \text{and} \quad \sum_{i=1}^m u_i E_i = 0. \quad (5.6)$$

Then the evolution law should have the form

$$u_i = -\Gamma \frac{\beta}{1 - \beta} \frac{N_i^{\beta-1}}{\sum_{i=1}^m N_i^\beta} + \lambda_1 E_i + \lambda_2, \quad (5.7)$$

where λ_1 and λ_2 are Lagrange multipliers that can be defined by substitution of (5.7) into the set of constraints (5.6).

The solutions for λ_1 and λ_2 are given by formulae

$$\left. \begin{aligned} \lambda_1 &= -\Gamma \frac{\beta}{1 - \beta} \frac{1}{\sum_{i=1}^m N_i^\beta} \frac{(\sum_{i=1}^m N_i^{\beta-1} \sum_{i=1}^m E_i - m \sum_{i=1}^m N_i^{\beta-1} E_i)}{m \sum_{i=1}^m E_i^2 - (\sum_{i=1}^m E_i)^2} \\ \text{and} \quad \lambda_2 &= -\Gamma \frac{\beta}{1 - \beta} \frac{1}{\sum_{i=1}^m N_i^\beta} \frac{(\sum_{i=1}^m N_i^{\beta-1} E_i \sum_{i=1}^m E_i - \sum_{i=1}^m N_i^{\beta-1} \sum_{i=1}^m E_i^2)}{m \sum_{i=1}^m E_i^2 - (\sum_{i=1}^m E_i)^2} \end{aligned} \right\} \quad (5.8)$$

This solution is defined for $(m \sum_{i=1}^m E_i^2 - (\sum_{i=1}^m E_i)^2) \neq 0$. It holds for all cases except a degenerate case when all E_i are equal.

The general form of the evolution law can be obtained by substitution of λ_1 and λ_2 from (5.8) into equation (5.7). In abbreviated form, we represent this law as

$$\dot{N} = -\Gamma \frac{\beta}{1-\beta} (I - A) N^{\beta-1}, \quad (5.9)$$

where $N = (N_1, \dots, N_m)^T$, I is an identity matrix and A is a symmetric $m \times m$ matrix defined as follows:

$$a_{ij} = \frac{1}{m} + \frac{\tilde{E}_i \tilde{E}_j}{\|\tilde{E}\|^2 - (1/m)(\mathbf{1}\tilde{E})^2},$$

where $\tilde{E}_i = E_i + (1/m) \sum_{i=1}^m E_i$, $\mathbf{1} = (1, \dots, 1)^T$ and $\tilde{E} = (E_1, \dots, E_m)^T$ is a vector of energies.

As before it can be shown that $\dot{V}(X, \beta) = H_{\max}(\beta) - H(X, \beta)$ is a Lyapunov function and there is only one stable equilibrium state of the system in non-degenerate cases. We demonstrate this as:

$$\dot{V}(\beta) = -\dot{H}(X, \beta) = \frac{\beta}{1-\beta} \frac{\sum_{i=1}^m N_i^{\beta-1} u_i}{\sum_{i=1}^m N_i^\beta}. \quad (5.10)$$

We substitute λ_1 and λ_2 from (5.8) into equation (5.7) and then we substitute the expression for u_i into (5.10). The resulting expression is

$$\begin{aligned} \dot{V}(\beta) = & \frac{\Gamma \beta^2}{(1-\beta)^2} \frac{1}{(\sum_{i=1}^m N_i^\beta)^2} \\ & \times \left(\frac{(m \sum_{i=1}^m E_i N_i^{\beta-1} - \sum_{i=1}^m E_i \sum_{i=1}^m N_i^{\beta-1})^2}{m \sum_{i=1}^m E_i^2 - (\sum_{i=1}^m E_i)^2} - \left(m \sum_{i=1}^m (N_i^{\beta-1})^2 - \left(\sum_{i=1}^m N_i^{\beta-1} \right)^2 \right) \right). \end{aligned} \quad (5.11)$$

We introduce a new scalar product function for two vectors as

$$\langle f, g \rangle = m \sum_i f_i g_i - \sum_i f_i \sum_i g_i. \quad (5.12)$$

For scalar product (5.12), the CBS inequality is true:

$$\langle f, g \rangle^2 \leq \langle f, f \rangle \langle g, g \rangle. \quad (5.13)$$

An equality in (5.13) takes place when $\exists \lambda, \mu \in \mathbb{R} : f = \lambda g + \mu$.

Using inequality (5.13) for vectors $f = (N_1^{\beta-1}, \dots, N_m^{\beta-1})^T$ and $g = (E_1, \dots, E_m)^T$, we get for (5.11) that $\dot{V}(\beta) \leq 0$. And $\dot{V}(\beta) = 0$ occurs only for the case when $\exists \lambda, \mu \in \mathbb{R} : N_i^{\beta-1} = \lambda E_i + \mu$ for all i . Due to (5.7) at the equilibrium state of the system, the following equalities hold:

$$-\Gamma \frac{\beta}{1-\beta} \frac{N_i^{\beta-1}}{\sum_{i=1}^m N_i^\beta} + \lambda_1 E_i + \lambda_2 = 0. \quad (5.14)$$

This means that, for the equilibrium state, the final distribution of points satisfies

$$N_i^{\beta-1} = \lambda E_i + \mu,$$

where $\lambda = \lambda_1((1-\beta)/\Gamma\beta) \sum_{i=1}^m N_i^\beta$, $\mu = \lambda_2((1-\beta)/\Gamma\beta) \sum_{i=1}^m N_i^\beta$ for λ_1 and λ_2 defined in (5.8).

(i) Final distribution and the Rényi distribution equivalence

Let us show that the distribution in (5.14) is the RD (4.3). If we multiply (5.14) by N_i and sum up over i , we obtain

$$-\Gamma \frac{\beta}{1-\beta} + \lambda_1 NE + \lambda_2 N = 0 \Rightarrow \lambda_2 = \frac{\Gamma' \beta}{1-\beta} - \lambda_1 E, \quad \Gamma' = \frac{\Gamma}{N}. \quad (5.15)$$

From (5.14), we get that

$$N_i = \left((\lambda_1 E_i + \lambda_2) \frac{1-\beta}{\Gamma \beta} \sum_{i=1}^m N_i^\beta \right)^{1/(\beta-1)}. \quad (5.16)$$

Let us substitute N_i from (5.16) into (5.1). We get

$$\begin{aligned} \frac{1}{N} \left(\frac{1-\beta}{\Gamma \beta} \sum_{i=1}^m N_i^\beta \right)^{1/(\beta-1)} \sum_{i=1}^m (\lambda_1 E_i + \lambda_2)^{1/(\beta-1)} &= 1 \\ \Rightarrow \left(\frac{1-\beta}{\Gamma \beta} \sum_{i=1}^m N_i^\beta \right)^{1/(\beta-1)} &= \frac{N}{\sum_{i=1}^m (\lambda_1 E_i + \lambda_2)^{1/(\beta-1)}}. \end{aligned} \quad (5.17)$$

After substitution of (5.17) into (5.16), we get that

$$\frac{N_i}{N} = \frac{(\lambda_1 E_i + \lambda_2)^{1/(\beta-1)}}{\sum_{i=1}^m (\lambda_1 E_i + \lambda_2)^{1/(\beta-1)}}. \quad (5.18)$$

Then we substitute the expression for λ_2 in (5.15) into (5.18) and obtain

$$\frac{N_i}{N} = \frac{(1 - \lambda((\beta-1)/\beta)(E_i - E))^{1/(\beta-1)}}{\sum_{i=1}^m (1 - \lambda((\beta-1)/\beta)(E_i - E))^{1/(\beta-1)}}, \quad (5.19)$$

where $\lambda = \lambda_1 / \Gamma'$. We can see that (5.5) coincides with the RD (5.19).

It is evident that (5.19) satisfies the normalization constraint (5.1). Let us check that the second constraint for energy (5.5) is also satisfied.

Let us substitute N_i from (5.16) into (5.5). Then we get

$$\begin{aligned} \frac{1}{N} \left(\frac{1-\beta}{\Gamma \beta} \sum_{i=1}^m N_i^\beta \right)^{1/(\beta-1)} \sum_{i=1}^m (\lambda_1 E_i + \lambda_2)^{1/(\beta-1)} E_i &= E \\ \Rightarrow \left(\frac{1-\beta}{\Gamma \beta} \sum_{i=1}^m N_i^\beta \right)^{1/(\beta-1)} &= \frac{NE}{\sum_{i=1}^m (\lambda_1 E_i + \lambda_2)^{1/(\beta-1)} E_i}. \end{aligned} \quad (5.20)$$

After substituting (5.20) into (5.16), we get that

$$\begin{aligned} N_i &= ((\lambda_1 E_i + \lambda_2)^{1/(\beta-1)}) \left(\frac{NE}{\sum_{i=1}^m (\lambda_1 E_i + \lambda_2)^{1/(\beta-1)} E_i} \right) \\ \Rightarrow E_i N_i &= \frac{(\lambda_1 E_i + \lambda_2)^{1/(\beta-1)} NE}{\sum_{i=1}^m (\lambda_1 E_i + \lambda_2)^{1/(\beta-1)} E_i} \Rightarrow \sum_{i=1}^m N_i E_i = NE, \end{aligned} \quad (5.21)$$

which means that energy constraint (5.5) is true for (5.19).

6. System with continuous probability density function

Let us extend the same approach based on the SG principle for the case of continuous PDFs. Consider a system with a continuous distribution of possible states that evolves on a compact

carrier. Distribution over states is characterized by PDF $p(t, x)$, which is continuous everywhere except for a set with zero measure. It is true that

$$\int_{\Omega} p(t, x) dx = 1, \quad \forall t, \quad (6.1)$$

where Ω is a compact carrier.

The Rényi entropy for a continuous PDF is defined as

$$H(X, \beta) = \frac{1}{1 - \beta} \log \int_{\Omega} p^{\beta}(t, x) dx, \quad (6.2)$$

where $\beta \geq 0, \beta \neq 1$.

From constraint (6.1), it follows that

$$\int_{\Omega} u(t, x) dx = 0, \quad (6.3)$$

where $u(t, x) = \dot{p}(t, x)$.

According to the SG principle, we calculate \dot{H} :

$$\dot{H}(X, \beta) = \frac{\beta}{1 - \beta} \frac{\int_{\Omega} \dot{p}(t, x) p^{\beta-1}(t, x) dx}{\int_{\Omega} p^{\beta}(t, x) dx}.$$

The gradient of \dot{H} by u is equal to

$$\nabla_u \dot{H}(X, \beta) = \frac{\beta}{1 - \beta} \nabla_u \left\langle \frac{p^{\beta-1}(t, x)}{\int_{\Omega} p^{\beta}(t, x) dx}, u \right\rangle = \frac{\beta}{1 - \beta} \frac{p^{\beta-1}(t, x)}{\int_{\Omega} p^{\beta}(t, x) dx}. \quad (6.4)$$

The SG principle of motion forms the evolution law:

$$u = -\Gamma \frac{\beta}{1 - \beta} \frac{p^{\beta-1}(t, x)}{\int_{\Omega} p^{\beta}(t, x) dx} + \lambda',$$

where Γ can be taken as a scalar value and the Lagrange multiplier λ' is selected to satisfy the constraint (6.3),

$$\int_{\Omega} \left(-\Gamma \frac{\beta}{1 - \beta} \frac{p^{\beta-1}(t, x)}{\int_{\Omega} p^{\beta}(t, x) dx} + \lambda' \right) dx = 0 \Rightarrow \lambda' = \frac{\beta \Gamma}{(1 - \beta) \text{mes}(\Omega)} \frac{\int_{\Omega} p^{\beta-1}(t, x) dx}{\int_{\Omega} p^{\beta}(t, x) dx}, \quad (6.5)$$

where $\text{mes}(\Omega) = \int_{\Omega} 1 dx$.

The final system dynamics equation has the following form:

$$\dot{p} = -\Gamma \frac{\beta}{1 - \beta} \frac{p^{\beta-1}(t, x)}{\int_{\Omega} p^{\beta}(t, x) dx} + \frac{\beta \Gamma}{(1 - \beta) \text{mes}(\Omega)} \frac{\int_{\Omega} p^{\beta-1}(t, x) dx}{\int_{\Omega} p^{\beta}(t, x) dx}. \quad (6.6)$$

Equation (6.6) can be represented in the more general form

$$\dot{p} = -\frac{\Gamma}{C} (I - \Psi) p^{\beta-1}, \quad (6.7)$$

where $\Psi = \int_{\Omega} (\cdot) dx / \text{mes}(\Omega)$ is a linear operator which is invariant for p , I is an identity operator and $C = (1 - \beta) \int_{\Omega} p^{\beta}(t, x) dx / \beta$.

(a) Equilibrium stability

Let us investigate a stability of the obtained equilibrium equation (6.6). Consider function $V(p) = H_{\max} - H(\beta) \geq 0$. The derivative of this function is

$$\dot{V}(p) = -\dot{H}(\beta) = -\frac{\beta}{1-\beta} \frac{\int_{\Omega} u(t, x) p^{\beta-1}(t, x) dx}{\int_{\Omega} p^{\beta}(t, x) dx}. \quad (6.8)$$

After substitution of the expression for u from (6.6) to (6.8), we obtain:

$$\dot{V}(p) = \frac{\Gamma \beta^2}{\text{mes}(\Omega)(\beta-1)^2 (\int_{\Omega} p^{\beta}(t, x) dx)^2} \left(\left(\int_{\Omega} p^{\beta-1}(t, x) dx \right)^2 - \text{mes}(\Omega) \int_{\Omega} (p^{\beta-1}(t, x))^2 dx \right). \quad (6.9)$$

Then we use the CBS inequality in integral form

$$\left| \int_{\Omega} f(x)g(x) dx \right|^2 \leq \left(\int_{\Omega} |f(x)|^2 \right) \left(\int_{\Omega} |g(x)|^2 \right) \quad (6.10)$$

for functions $f = p^{\beta-1}$ and $g = 1$. Taking into account that a scalar value Γ is positive, we get that $\dot{V}(p) \leq 0$. It is known that equality in the CBS inequality is achieved when multiplicity occurs, i.e. $f(x) = \alpha g(x)$. In our case, $\dot{V}(p) = 0$ is true when $p^{\beta-1} = \alpha$. This is possible only when $p(t, x) = C = \text{const}$. Using the constraint (6.1), we get that $C = \text{mes}^{-1}(\Omega)$. This means that there is only one unique limit PDF $p^* = \text{mes}^{-1}(\Omega)$ for the equilibrium state of the system which evolves by evolution law (6.6). Also note that p^* corresponds to RD (4.4) for a case with one constraint.

(b) Asymptotic convergence

To show an asymptotic convergence of all solutions to p^* , we will use Barbalat's lemma [36].

Lemma 6.1 (Barbalat's lemma). *If differentiable function $f(t)$ has a finite limit for $t \rightarrow \infty$ and its derivative $\dot{f}(t)$ is uniformly continuous then $\dot{f}(t) \rightarrow 0$ for $t \rightarrow \infty$.*

Theorem 6.2. *For all PDFs defined by equation (6.6), it is true that $p(t, x) \rightarrow p^*$ for $t \rightarrow \infty$.*

Proof. For the sake of simplicity, we define a notation for V in (6.9) as $v(t) = V(p(t))$. We will use Barbalat's lemma to show that $\dot{v}(t) \rightarrow 0$. We use $v(t)$ as a function $f(t)$ in Barbalat's lemma. Because of $v(t) \geq 0$ and $\dot{v} \leq 0$, the function $v(t)$ has a finite limit for $t \rightarrow \infty$.

It can be shown that \dot{v} is uniformly continuous. Consider the expression for $|\ddot{v}(t)|$,

$$\begin{aligned} \ddot{v}(t) = & \frac{\beta}{\beta-1} \frac{\int_{\Omega} (\dot{u}(t, x) p^{\beta-1}(t, x) + (\beta-1) u^2(t, x) p^{\beta-2}(t, x)) dx}{\int_{\Omega} p^{\beta} dx} \\ & - \frac{\beta^2}{\beta-1} \int_{\Omega} u(t, x) p^{\beta-1}(t, x) dx \frac{\int_{\Omega} p^{\beta-1}(t, x) u(t, x) dx}{(\int_{\Omega} p^{\beta}(t, x) dx)^2}. \end{aligned} \quad (6.11)$$

Due to the constraint (6.1) and compactness of the carrier Ω , it can be shown that function $|\ddot{v}(t)|$ is bounded. This leads us to the fact that \dot{v} is uniformly continuous.

As all the necessary conditions of Barbalat's lemma for differentiable function $v(t)$ are satisfied, we have that $\dot{v}(t) \rightarrow 0$ for $t \rightarrow \infty$.

Taking into account that $\int_{\Omega} f dx = (1, f)$ and $\int_{\Omega} f^2 dx = \|f\|^2$, the expression for \dot{v} from (6.9) may be rewritten as

$$\begin{aligned} \dot{v} = & -\frac{\Gamma \beta^2}{\text{mes}(\Omega)(\beta-1)^2} \frac{1}{(p^{\beta}, 1)^2} \left(1 - \frac{(1, p^{\beta-1})^2}{\text{mes}(\Omega) \|p^{\beta-1}\|^2} \right) \\ = & -\frac{\Gamma \beta^2}{\text{mes}(\Omega)(\beta-1)^2} \frac{\|p^{\beta-1}\|^2}{(p^{\beta}, 1)^2} (1 - \cos^2(\alpha(t))), \end{aligned} \quad (6.12)$$

where $\alpha(t) = \cos(\langle 1, p^{\beta-1} \rangle)$.

If $\|p^{\beta-1}\|^2 \rightarrow 0$, then $\text{mes}_t\{r: p_t(r) \neq 0\} \rightarrow 0$. But this case conflicts with the constraint (6.1). Given (6.2) we obtain that $\alpha(t) \rightarrow 0$. This means that $\hat{p}^{\beta-1} \rightarrow \hat{1}$, where $\hat{p}^{\beta-1}$ and $\hat{1}$ are normalized values for $p^{\beta-1}$ and 1, respectively.

It follows that $p(t, x)$ tends to the stationary distribution. As explained earlier, this distribution is unique. Thus, $p_t \rightarrow p^*$ for $t \rightarrow \infty$. ■

(c) Total energy constraint

The constraint (6.1) can be interpreted as the mass conservation law on the space Ω . Consider a system with an additional constraint for the total energy conservation, i.e. a conservative case when energy does not depend on time. The new constraint may be described as

$$\int_{\Omega} p(t, x) h(x) dx = E, \quad (6.13)$$

where E is the common energy of a system and $h(x)$ is the density of energy.

The equation for the dynamics can be defined in the form

$$u = -\Gamma \nabla_u \dot{H}(\beta) + \lambda_1 h + \lambda_2. \quad (6.14)$$

Based on constraints (6.1) and (6.13), we can find expressions for Lagrange multipliers λ_1 and λ_2 :

$$\left. \begin{aligned} \lambda_1 &= \frac{\Gamma \beta}{(1 - \beta) \int_{\Omega} p^{\beta}(t, x) dx} \frac{\text{mes}(\Omega) \int_{\Omega} p^{\beta-1}(t, x) h(x) dx - \int_{\Omega} p^{\beta-1}(t, x) dx \int_{\Omega} h(x) dx}{\text{mes}(\Omega) \int_{\Omega} h(x)^2 dx - (\int_{\Omega} h(x) dx)^2} \\ \text{and } \lambda_2 &= \frac{\Gamma \beta}{(1 - \beta) \int_{\Omega} p^{\beta}(t, x) dx} \frac{\int_{\Omega} p^{\beta-1}(t, x) dx \int_{\Omega} h^2(x) dx - \int_{\Omega} p^{\beta-1}(t, x) h(x) dx \int_{\Omega} h(x) dx}{\text{mes}(\Omega) \int_{\Omega} h(x)^2 dx - (\int_{\Omega} h(x) dx)^2} \end{aligned} \right\} \quad (6.15)$$

The above equations are valid when the denominator in both fractions is not equal to zero. If we use the CBS inequality for $f = h$ and $g = 1$, then the following inequality becomes true: $|\int_{\Omega} h dx|^2 \leq \text{mes}(\Omega) \int_{\Omega} h^2 dx$. This inequality becomes equality when $h = \text{const}$. This means that all energy levels coincide. This case is supposed to be degenerate and not considered here. Thus the expression $\text{mes}(\Omega) \int_{\Omega} h^2 dx \neq (\int_{\Omega} h dx)^2$ is always true.

The resulting equation for the dynamics can be obtained by substituting (6.4) and (6.15) into (6.14). It can be transformed to the brief form:

$$\dot{p} = -\frac{\Gamma}{C} (I - \Psi) p^{\beta-1}, \quad (6.16)$$

where I is an identity operator and Ψ is a linear integral operator that is independent of p :

$$\Psi = \frac{(1, \cdot)}{\text{mes}(\Omega)} + \frac{\tilde{h}(\tilde{h}, \cdot)}{\|h\|^2 - \frac{1}{\text{mes}(\Omega)}(1, h)^2}, \quad (6.17)$$

where $\tilde{h} = h - (1/\text{mes}(\Omega)) \int_{\Omega} h dx$, $C = (1 - \beta) \int_{\Omega} p^{\beta}(t, x) dx / \beta$.

(i) Equilibrium stability

Let us examine the equilibrium of the obtained equation (6.14). We use the same Lyapunov function \dot{V} as we have used in the previous section with only one constraint. For two constraints, the new expression for \dot{V} is

$$\begin{aligned} \dot{V}(p) &= \frac{\Gamma \beta^2}{\text{mes}(\Omega)(\beta - 1)^2 (\int_{\Omega} p^{\beta}(t, x) dx)^2} \frac{(\text{mes}(\Omega) \int_{\Omega} h p^{\beta-1} dx - \int_{\Omega} h dx \int_{\Omega} (p^{\beta-1})^2 dx)^2}{\text{mes}(\Omega) \int_{\Omega} h^2 dx - (\int_{\Omega} h dx)^2} \\ &\quad - \frac{\Gamma \beta^2}{\text{mes}(\Omega)(\beta - 1)^2 (\int_{\Omega} p^{\beta}(t, x) dx)^2} \left(\text{mes}(\Omega) \int_{\Omega} (p^{\beta-1}(t, x))^2 dx - \left(\int_{\Omega} p^{\beta-1}(t, x) dx \right)^2 \right). \end{aligned} \quad (6.18)$$

We will prove that

$$\dot{V}(p) \leq 0. \quad (6.19)$$

Let us define a functional: $\langle \cdot, \cdot \rangle : L_2(\Omega) \times L_2(\Omega) \rightarrow \mathbb{R}$, $\forall f, g \in L_2(\Omega)$

$$\langle f, g \rangle = \text{mes}(\Omega) \int_{\Omega} fg \, dx - \int_{\Omega} f \, dx \int_{\Omega} g \, dx. \quad (6.20)$$

The new functional has several useful properties (the proof of each property is provided in appendix A):

- (1) linearity for the first argument $\forall f, g, h \in L_2(\Omega)$, $\forall \lambda \in \mathbb{R}$ $\langle \lambda f + g, h \rangle = \langle \lambda f, h \rangle + \langle g, h \rangle$;
- (2) symmetry $\forall f, g \in L_2(\Omega)$ $\langle f, g \rangle = \langle g, f \rangle$;
- (3) positiveness and the condition of zero value $\forall f \in L_2(\Omega)$ $\langle f, f \rangle \geq 0$, $\langle f, f \rangle = 0 \Leftrightarrow f = \mu = \text{const.}$

Let us prove inequality (6.19) based on properties 1–3.

It is obvious that for any $f, g \in L_2(\Omega)$ and $\lambda \in \mathbb{R}$ it is true that $f - \lambda g \in L_2(\Omega)$. This function has property 3: $\langle f - \lambda g, f - \lambda g \rangle \geq 0$. Using properties 1 and 2, we obtain the quadratic inequality with respect to λ :

$$0 \leq \langle f - \lambda g, f - \lambda g \rangle = \lambda^2 \langle g, g \rangle - 2\lambda \langle f, g \rangle + \langle f, f \rangle.$$

This inequality holds for any real λ . Hence the discriminant cannot be positive,

$$D = 4\langle f, g \rangle^2 - 4\langle f, f \rangle \langle g, g \rangle.$$

Thus

$$\langle f, g \rangle^2 \leq \langle f, f \rangle \langle g, g \rangle. \quad (6.21)$$

If the equality takes place in (6.21), then there exists a unique solution $\lambda \in \mathbb{R}$ of equation $\langle f - \lambda g, f - \lambda g \rangle = 0$. But then, by property 3, we have $\exists \mu \in \mathbb{R} : f - \lambda g = \mu \mathbf{1}$.

Substituting $f = p^{\beta-1}$, $g = h$ into inequality (6.21), we get

$$\begin{aligned} & \left(\text{mes}(\Omega) \int_{\Omega} h p^{\beta-1} \, dx - \int_{\Omega} h \, dx \int_{\Omega} p^{\beta-1} \, dx \right)^2 \\ & \leq \left(\text{mes}(\Omega) \int_{\Omega} (p^{\beta-1})^2 p \, dx - \left(\int_{\Omega} p^{\beta-1} \, dx \right)^2 \right) \left(\text{mes}(\Omega) \int_{\Omega} h^2 p \, dx - \left(\int_{\Omega} h \, dx \right)^2 \right), \end{aligned}$$

which implies inequality (6.19).

Note that equality (6.19) holds if and only if

$$\exists \lambda, \mu \in \mathbb{R} : p^{\beta-1} = \lambda h + \mu. \quad (6.22)$$

According to the SG for H (6.4), expression (6.14) can be rewritten for the case of equilibrium as

$$p(t, x)^{\beta-1} = \frac{C\lambda_1 h(x)}{\Gamma} + \frac{C\lambda_2}{\Gamma}, \quad (6.23)$$

where $C = (1 - \beta) \int_{\Omega} p^{\beta}(t, x) \, dx / \beta$. This coincides with (6.22) for $\lambda = \lambda_1(C/\Gamma)$, $\mu = \lambda_2(C/\Gamma)$, where λ_1 and λ_2 are defined in (6.15). So this is the only PDF $p^*(\beta)$ for an equilibrium state.

It can also be shown that $p^*(\beta)$ corresponds to the continuous form of the RD (4.4). To do this, we write expression (6.23) as

$$p(t, x) = \left((\lambda_1 h(x) + \lambda_2) \frac{1}{\Gamma} \frac{(1 - \beta)}{\beta} \int_{\Omega} p^{\beta}(t, x) \, dx \right)^{1/(\beta-1)}. \quad (6.24)$$

Following a similar approach to that used for a discrete case, expression (6.24) can be formulated as RD:

$$p(t, x) = \frac{(1 - \lambda((\beta - 1)/\beta)(h(x) - E))^{1/(\beta-1)}}{\int_{\Omega} (1 - \lambda((\beta - 1)/\beta)(h(x) - E))^{1/(\beta-1)} dx},$$

where $\lambda = \lambda_1/\Gamma$.

(ii) Asymptotic convergence

We will prove asymptotic convergence similarly to the case with one constraint (see proposition 1).

Theorem 6.3. *For all PDFs defined by equation (6.14), it is true that $p(t, x) \rightarrow p^*$ for $t \rightarrow \infty$.*

Proof. To use Barbalat's lemma, we have to check the conditions under which the function \dot{v} is bounded. Based on the expression for \dot{v} in (6.18) and according to similar logic so that used in proposition 1, we can conclude that \dot{v} is bounded for the compact carrier Ω . According to Barbalat's lemma, it is true that

$$\dot{v} \rightarrow 0, \quad t \rightarrow \infty. \quad (6.25)$$

We introduce a scalar product as $\langle f, f \rangle = \text{mes}(\Omega) \int_{\Omega} f^2(x) dx - (\int_{\Omega} f(x) dx)^2$. Having $\|f\|^2 = \langle f, f \rangle$, expression (6.18) can be rewritten as

$$\begin{aligned} \dot{v} &= \frac{\Gamma \beta^2}{\text{mes}(\Omega)(\beta - 1)^2 (\int_{\Omega} p^{\beta}(t, x) dx)^2} \left(\|p^{\beta-1}\|^2 - \frac{(p^{\beta-1}, h)^2}{\|h\|^2} \right) \\ &= \frac{\Gamma \beta^2 \|p^{\beta-1}\|^2}{\text{mes}(\Omega)(\beta - 1)^2 (\int_{\Omega} p^{\beta}(t, x) dx)^2} \left(1 - \frac{(p^{\beta-1}, h)^2}{\|p^{\beta-1}\|^2 \|h\|^2} \right). \end{aligned} \quad (6.26)$$

As $\dot{v} \rightarrow 0$, consider the case when $\|p^{\beta-1}\|^2 \rightarrow 0$ first. In the CBS inequality

$$\left(\int_{\Omega} p^{\beta-1} dx \right)^2 \leq \text{mes}(\Omega) \int_{\Omega} (p^{\beta-1})^2 p dx,$$

the equality takes place only when $p^{\beta-1} = \alpha$. We have previously demonstrated in (6.23) that equality $\dot{v} = 0$ holds only in the case when $p^{\beta-1} = \lambda h + \mu$, where λ and μ are constants. Then h must be a constant since the equality $\lambda h + \mu = \alpha$. We assume this to be a degenerate case.

Given (6.25), (6.26) and $\|p^{\beta-1}\|^2 \not\rightarrow 0$, we have that $(p^{\beta-1}, h)^2 / \|p^{\beta-1}\|^2 \|h\|^2 \rightarrow 1$. This implies that $\hat{p}^{\beta-1} \rightarrow \hat{h}$, where $\hat{p}^{\beta-1}$ and \hat{h} are normalized values for $p^{\beta-1}$ and h , respectively. Thus p tends to the only stationary distribution $p^*(\beta)$ since h does not depend on time. ■

7. Conclusion

The Rényi entropy is widely used in communication and coding theory, quantum information theory, signal processing, data mining and many other areas [4,5]. Stationary states which maximize Rényi entropy have already been well investigated. The MaxEnt principle defines the asymptotic behaviour of the system, but does not answer the question about how the system moves to this asymptotic behaviour.

We have investigated non-stationary states of processes that follow the MaxEnt principle for the Rényi entropy. We have derived equations (5.3), (5.9), (6.6) and (6.16) which describe the dynamics of the PDF for the system that tends to the state with maximum Rényi entropy. Systems with discrete probability distribution and continuous PDFs were considered under mass conservation and energy conservation constraints. We have shown that the limit PDF $p^*(\beta)$ is unique and corresponds to the RD. We have also proved the convergence of PDFs with dynamics described by equations (5.3), (5.9), (6.6) and (6.16) to PDF $p^*(\beta)$, which corresponds to the state with the maximum value of the Rényi entropy.

The key point of our approach is to use the SG method with the goal function chosen as the Rényi entropy of the process. The SG principle originates from control theory and it generates

equations for the transient (non-stationary) states of the system's operation which help to track how the system evolves to the steady state.

There are many other generalizations of the Shannon entropy, e.g. the Tsallis entropy [37], the Khinchin entropy [38], the Burbea–Rao entropy [39], the Cressie–Read family [40], the Burg entropy [41], f-divergences [31], etc. Investigation of these entropies based on the SG principle seems to be promising for further investigations. A comparative analysis of the dynamics for different kinds of entropy may be performed.

Data accessibility. All methods and corresponding data are given in the manuscript and are fully reproducible.

Authors' contributions. A.L.F. designed the study to use the SG principle for non-stationary processes which tend to maximize its entropy. D.S.S. derived the equations for the Rényi entropy and prepared the manuscript. Both authors gave final approval for publication.

Competing interests. We have no competing interests.

Funding. This work was supported by SPbU grant nos. 6.37.181.2014 and 6.38.230.2015. Formulation of the SG principle for the Rényi entropy (§5) was performed in IPME RAS and supported by the RSF (grant no. 14-29-00142).

Appendix A. Extra materials

Here, we prove the properties of functional $\langle f, g \rangle$ (6.20) from §5c(i).

(1) Linearity in the first argument $\forall f, g, h \in L_2(\Omega)$, $\forall \lambda \in \mathbb{R}$ $\langle \lambda f + g, h \rangle = \langle \lambda f, h \rangle + \langle g, h \rangle$.

Proof. Using the linearity of the integral, we obtain

$$\begin{aligned} \langle \lambda f + g, h \rangle &= \text{mes}(\Omega) \int_{\Omega} (\lambda f + g)h \, dx - \int_{\Omega} (\lambda f + g) \, dx \int_{\Omega} h \, dx \\ &= \left(\text{mes}(\Omega) \int_{\Omega} \lambda f h \, dx - \int_{\Omega} \lambda f \, dx \int_{\Omega} h \, dx \right) + \left(\text{mes}(\Omega) \int_{\Omega} g h \, dx - \int_{\Omega} g \, dx \int_{\Omega} h \, dx \right) \\ &= \langle \lambda f, h \rangle + \langle g, h \rangle. \end{aligned}$$

(2) Symmetry $\forall f, g \in L_2(\Omega)$ $\langle f, g \rangle = \langle g, f \rangle$.

Proof.

$$\langle f, g \rangle = \text{mes}(\Omega) \int_{\Omega} f g \, dx - \int_{\Omega} f \, dx \int_{\Omega} g \, dx = \text{mes}(\Omega) \int_{\Omega} g f \, dx - \int_{\Omega} g \, dx \int_{\Omega} f \, dx = \langle g, f \rangle.$$

(3) Positiveness and the condition of zero value $\forall f \in L_2(\Omega)$ $\langle f, f \rangle \geq 0$, $\langle f, f \rangle = 0 \Leftrightarrow f = \mu = \text{const}$.

Proof. Let us consider a scalar multiplication $\langle f, g \rangle = \int_{\Omega} f(x)g(x) \, d\Omega$. The CBS inequality becomes true for: $|\langle f, g \rangle|^2 \leq \langle f, f \rangle \langle g, g \rangle$, $|\langle f, g \rangle|^2 = \langle f, f \rangle \langle g, g \rangle \Leftrightarrow \exists \mu \in \mathbb{R} : f = \mu g$. Substituting $g \equiv 1$, we get $|\int_{\Omega} f \, dx|^2 \leq \langle f, f \rangle \int_{\Omega} 1 \, dx = \text{mes}(\Omega) \int_{\Omega} f^2 \, dx$. Thereby, it is true that $\langle f, f \rangle = \text{mes}(\Omega) \int_{\Omega} f^2 \, dx - (\int_{\Omega} f \, dx)^2 \geq 0$. Moreover, $\langle f, f \rangle = 0 \Leftrightarrow |\int_{\Omega} f \, dx|^2 = \langle f, f \rangle \int_{\Omega} 1 \, dx \Leftrightarrow \exists \mu \in \mathbb{R} : f = \mu 1$, i.e. $f = \mu$.

References

1. Martyushev L, Seleznev V. 2006 Maximum entropy production principle in physics, chemistry and biology. *Phys. Rep.* **426**, 1–45. (doi:10.1016/j.physrep.2005.12.001)
2. Shannon C. 1948 A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423. (doi:10.1002/j.1538-7305.1948.tb01338.x)
3. Rényi A. 1961 On measures of entropy and information. *Proc. Fourth Berkeley Symp. Math. Stat. Probab.* **1**, 547–561.
4. Wilde MM. 2015 Recoverability in quantum information theory. *Proc. R. Soc. A* **471**, 20150338. (doi:10.1098/rspa.2015.0338)
5. Bercher J-F. 2008 On some entropy functionals derived from Rényi information divergence. *Inform. Sci.* **178**, 2489–2506. (doi:10.1016/j.ins.2008.02.003)

6. Jaynes E. 1980 The minimum entropy production principle. *Annu. Rev. Phys. Chem.* **31**, 579–601. (doi:10.1146/annurev.pc.31.100180.003051)
7. Jaynes E. 1957 Information theory and statistical mechanics I. *Phys. Rev.* **106**, 620–630. (doi:10.1103/PhysRev.106.620)
8. Jaynes E. 1957 Information theory and statistical mechanics II. *Phys. Rev.* **108**, 171–190. (doi:10.1103/PhysRev.108.171)
9. Ortega PA, Braun DA. 2013 Thermodynamics as a theory of decision-making with information-processing costs. *Proc. R. Soc. A* **469**, 20120683. (doi:10.1098/rspa.2012.0683)
10. Lucia U. 2008 Probability, ergodicity, irreversibility and dynamical systems. *Proc. R. Soc. A* **464**, 1089–1104. (doi:10.1098/rspa.2007.0304)
11. Martyushev LM. 2013 Entropy and entropy production: old misconceptions and new breakthroughs. *Entropy* **15**, 1152–1170. (doi:10.3390/e15041152)
12. Nagy Á, Romera E. 2009 Maximum Rényi entropy principle and the generalized Thomas–Fermi model. *Phys. Lett. A* **373**, 844–846. (doi:10.1016/j.physleta.2009.01.004)
13. Fradkov AL. 2008 Speed-gradient entropy principle for nonstationary processes. *Entropy* **10**, 757–764. (doi:10.3390/e10040757)
14. Fradkov A, Miroshnik I, Nikiforov V. 1999 *Nonlinear and adaptive control of complex systems*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
15. Fradkov A, Krivtsov A. 2011 Speed-gradient principle for description of transient dynamics in systems obeying maximum entropy principle. In *Proc. of the 30th Int. Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Chamonix, France, 4–9 July 2010*. AIP Conference Proceedings, vol. 1305, pp. 399–406. College Park, MD: American Institute of Physics.
16. Fradkov A. 1979 Speed-gradient scheme and its application in adaptive-control problems. *Autom. Remote Control* **40**, 1333–1342.
17. Fradkov AL. 2007 *Cybernetical physics: from control of chaos to quantum control*. Berlin, Germany: Springer.
18. Fradkov AL, Shalymov DS. 2015 Dynamics of non-stationary nonlinear processes that follow the maximum of differential entropy principle. *Commun. Nonlinear Sci. Numer. Simul.* **29**, 488–498. (doi:10.1016/j.cnsns.2015.06.001)
19. Fradkov AL, Shalymov DS. 2015 Speed-gradient and MaxEnt principles for Shannon and Tsallis entropies. *Entropy* **17**, 1090–1102. (doi:10.3390/e17031090)
20. Frank TD. 2005 *Nonlinear Fokker–Planck equations*. Berlin, Germany: Springer.
21. Hick P, Stevens G. 1987 Minimum Kullback entropy approach to the Fokker–Planck equation. *Astron. Astrophys.* **172**, 350.
22. Plastino AR, Miller HG, Plastino A. 1997 Minimum Kullback entropy approach to the Fokker–Planck equation. *Phys. Rev. E* **56**, 3927. (doi:10.1103/PhysRevE.56.3927)
23. Grmela M, Öttinger H. 1997 Dynamics and thermodynamics of complex fluids. I. Development of a general formalism. *Phys. Rev. E* **56**, 6620. (doi:10.1103/PhysRevE.56.6620)
24. Öttinger H, Grmela M. 1997 Dynamics and thermodynamics of complex fluids. II. Illustrations of a general formalism. *Phys. Rev. E* **56**, 6633. (doi:10.1103/PhysRevE.56.6633)
25. Öttinger H. 1998 General projection operator formalism for the dynamics and thermodynamics of complex fluids. *Phys. Rev. E* **57**, 1416–1420. (doi:10.1103/PhysRevE.57.1416)
26. Ho S-W, Yeung RW. 2010 The interplay between entropy and variational distance. *IEEE Trans. Inform. Theory* **56**, 5906–5929. (doi:10.1109/TIT.2010.2080452)
27. Ho S-W, Yeung RW. 2010 On information divergence measures and a unified typicality. *IEEE Trans. Inform. Theory* **56**, 5893–5905. (doi:10.1109/TIT.2010.2080431)
28. Shore JE, Johnson RW. 1980 Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inform. Theory* **26**, 26–37. (doi:10.1109/TIT.1980.1056144)
29. Ali SA, Cafaro C, Giffin A, Kim DH. 2012 Complexity characterization in a probabilistic approach to dynamical systems through information geometry and inductive inference. *Phys. Scr.* **85**, 025009. (doi:10.1088/0031-8949/85/02/025009)
30. Cafaro C. 2008 Information geometry, inference methods and chaotic energy levels statistics. *Mod. Phys. Lett. B* **22**, 1879–1892. (doi:10.1142/S0217984908016558)
31. Morimoto T. 1963 Markov processes and the h-theorem. *J. Phys. Soc. Jpn* **12**, 328–331.
32. Lanczos K. 1962 *The variational principles of mechanics*. Toronto, Canada: University of Toronto Press.

33. Planck M. 1909 The unity of physical world views. *Physik. Z.* **10**, 62–75.
34. Bashkirov AG. 2004 Maximum Rényi entropy principle for systems with power-law Hamiltonians. *Phys. Rev. Lett.* **93**, 130601. (doi:10.1103/PhysRevLett.93.130601)
35. Bashkirov AG. 2004 On maximum entropy principle, superstatistics, power-law distribution and Rényi parameter. *Physica A* **340**, 153–162. (doi:10.1016/j.physa.2004.04.002)
36. Lyapunov AM. 1966 *Stability of motion*. New York, NY: Academic Press.
37. Tsallis C. 1988 Possible generalization of Boltzmann–Gibbs statistics. *J. Stat. Phys.* **52**, 479–487. (doi:10.1007/BF01016429)
38. Khinchin A. 1957 *Mathematical foundation of information theory*. New York, NY: Dover.
39. Burbea J, Rao C. 1982 On the convexity of some divergence measures based on entropy functions. *IEEE Trans. Inform. Theory* **28**, 489–495. (doi:10.1109/TIT.1982.1056497)
40. Cressie N, Read T. 1984 Multinomial goodness of fit tests. *J. R. Stat. Soc. Ser. B* **46**, 440–464.
41. Burg J. 1972 The relationship between maximum entropy spectra and maximum likelihood spectra. *Geophysics* **37**, 375–376. (doi:10.1190/1.1440265)