Pairwise Interactions Origin of Entropy Functions

Yuri A. Pykh

Research Center for Interdisciplinary Environmental Cooperation of Russian Academy of Sciences (INENCO RAS), Saint-Petersburg, 191187, nab. Kutuzova, 14, Russia; Email: inenco@mail.neva.ru, malkinapykh@yandex.ru; Tel: +7 – 921 – 9034092

Abstract: In this paper we combine the three universalisms: pairwise interactions concept, dynamical systems theory and relative entropy analysis to develop a theory of entropy issues. We introduce two hypotheses concerning the structure and types properties of the system’s entities interactions and derive generalized replicator dynamic equations. Then we construct energy-like and entropy-like Lyapunov-Meyer functions (LMF) for these equations. We show that energy-like LMF contains no information about the equilibrium of the system and is a substantial generalization of the Fisher’s fundamental theorem of natural selection. If there are exists nontrivial equilibrium point for generalized replicator system then we construct entropy-like LMF for this system and prove that it is a relative entropy function or the function of information divergence. We prove that negative relative entropy is a convex function for a probability space and receive new distance measure between two probability distributions. Also we use Legendre-Donkin-Fenchel transformation for dual coordinates. As result we establish the set of next important links: nonlinear pairwise interaction → generalized Fisher (replicator) equations → Lyapunov-Meyer functions → relative entropy → distance measure → LDF-transformation → duality.

In particular it follows from these links that nonlinear pairwise interaction is the origin of all known entropy functions.

Keywords: pairwise interactions; Lyapunov functions; stability postulate; entropy functions; escort distribution
1. Introduction

1.1. Entropy and Lyapunov functions

The primary goal of this paper is to study the implications of Lyapunov functions theory for the analysis of nonlinear pairwise interactions replicator systems. We build two types of Lyapunov-Meyer functions, either as an energy-like or as an entropy-like functions (distance measures). The existence of the Lyapunov-Meyer functions allows us to construct practically all known entropy and relative-entropy functions.

Our investigations are based mainly on the stability postulate stated by Chetaev in 1936 [1].

**Stability postulate.** *Stability, which is a fundamentally general phenomenon, apparently, must manifest itself in basic laws of nature in some way. If knowledge is constructed from the requirement of small deviations from nature, then scientific thinking must (or can) rely on some Lyapunov function.*

Certainly, this function always exists according to the stability postulate.

Many years later in 1967 stability postulate, in more narrow sense, was used by B.D.Coleman and V.J. Mizel in famous paper “Existence of Entropy as a Consequence of Asymptotic stability” [2].

In 1968, this postulate was stated mathematically by Meyer [3], who proved that, for dynamical systems whose limit sets consist of only isolated rest points or cycles, i.e., for Morse—Smale systems, there always exists a Lyapunov function, which increases (or decreases) on the set of wandering points of the system. Meyer following by S.Smale [4] suggested the term energy functions for such functions; however, it is more natural and convenient to refer to them as Lyapunov–Meyer functions. One of the best known examples of such function is entropy in Boltzmann’s H-theorem.

Ilya Prigogine first pointed out [5] the importance of the relationship between Lyapunov functions and entropy: “The positive time direction is associated with the increase of entropy. Let us emphasize the strong and very specific way in which the one-sidedness of time appears in the second law. According to its formulation it leads to the existence of a function having quite specific properties as expressed by the fact that for an isolated system it can only increase in time. Such function plays an important role in modern theory of stability as initiated by the classic work of Lyapunov. For this reason they are called Lyapunov functions (or functionals).

The entropy S is a Lyapunov function for isolated systems. As shown in all textbooks thermodynamic potentials such as the Helmholtz or Gibbs free energy are also Lyapunov functions for other “boundary conditions” (such as imposed values of temperature and volume).

In all these cases the system evolves to an equilibrium state characterized by the existence of a thermodynamic potential. This equilibrium state is an “attractor” for non-equilibrium states.”
For dynamical systems arising from physics the Lyapunov functions will typically have a thermodynamic interpretation (energy, entropy etc.) but its origin is not evident.

In an attempt to show how such an interpretation this ideas might proceed, we analyze below a class of macrosystems with nonlinear pairwise interactions for with an assumption of structure and type of interactions has consequences that there are exist an entropy-like and energy-like LMF functions.

1.2. Pairwise Interactions in nature and science: a review and outlook:

In Bialek and Ranganatan's recent paper: "Rediscovering the power of pairwise interactions" [6] is emphasized that the analysis of many objects the considered in physics, chemistry, biology, economy and other natural sciences can be carried out on the basis of research of the interactions. The mathematical background of this idea was developed in the works [7,8] where on the basis of the well-known theorem of Kolmogorov-Arnold of representation of any continuous function by superposition of functions of one or two variables, showed that pairwise interactions possess the fundamental properties defining dynamics of processes in many physical tasks. In particular V. Kreinovich [8] have shown that: “An arbitrary particle interaction can be represented in pairwise form by adding a finite number of extra particles.” Supporting this general expectation we will carry out the small chronological review of the most known models of the pairwise interactions.

- **Newton's law of gravitation.** Most physical theories such as gravity or thermodynamics are based on pairwise interactions between particles. Thus, for example, the law of universal gravitation opened by I. Newton in 1684. According this law a force acting on the i-th body is equal to the sum of the corresponding pairwise forces. It is necessary to emphasize that this approach known under the name gravitational, is widely used at creation of models of dynamics of the most various interacting objects now [9].

- **Boltzmann’s kinetic equation.** Boltzmann's equation – the main equation in the kinetic theory of gases (1872). It describes change of distribution function of molecules on speeds and on coordinates. In approach of a hypothesis of pairwise collisions this equation is reduced to one closed equation for one-partial function of distribution – to the kinetic equation of Boltzmann. For this equation Boltzmann proves well-known H – theorem. Using the modern language it is possible to tell that he found for this equation a Lyapunov function related to entropy. We would note also that using Boltzmann's ideas in 1963 T. Hill [10], constructed thermodynamics of "small" systems which received further the name of nanothermodynamics [11].

- **Equation of Lotka-Volterra.** In 1920 A.Lotka [12] at the description of chemical reactions, kinetics and then in 1926 V. Volterra [13] for modeling dynamics of the interacting groups of biological community used a hypothesis of linear pairwise interactions that led to creation of well-
known system of the ordinary differential equations with square non-linearities. For these equations two types of Lyapunov functions were found: of entropy-like type and energy-like functions. Recently, A. Wilson [14] established connection between the kinetic equations of Boltzmann and Lotka-Volterra system. In the monograph [15] author analyzes generalized Lotka-Volterra system with nonlinear pairwise interactions, the paper [16] provides a survey of recent activity in this issue.

- **Izing Model.** In 1924 E.Izing [17] offered the mathematical model of ferromagnetism in statistical mechanics based on a hypothesis of pairwise interactions between spins. Now this model of statistical mechanics to which research the huge number of publications is devoted found application in many areas of natural sciences [18].

- **Fischer's equations.** In 1930 R. Fischer [19], developing the mathematical theory of populations genetic structure evolution, offered the equations of dynamics of this structure. This system of the equations, also as well as Lotka-Volterra system, is founded on a hypothesis of linear pairwise interactions that is quite natural to the individuals breeding in population. For this system Fischer formulated based on enegy-like Lyapunov's function the evolutionary principle known as the main theorem of natural selection. Later, in 1972, Yu.A.Pykh [20] constructs for this system Lyapunov's entropy-like function. The perfect analysis of classical Fischer’s equations with linear pairwise interactions is given in the monograph [15].

- **Equations of biological macromolecules evolution.** In 1971 M. Eigen [21] used the principle of pairwise interactions at creation the model of prebiological evolution explaining emergence of the ordered structures from the disorder. For a catalytic cycle with the constant general organization this approach gives the system of the equations close to Fischer's equations [15].

- **Evolutionary games.** In 1978 Taylor and Jonker [22] based on the concept of Evolutionarily Stable Strategy offered by Maynard Smith and using the principle of pairwise interactions constructed system of the equations the defining dynamics of the interacting populations depending on the strategy chosen by individuals. In this work for the first time there was using a name the replicator equations, entered in 1976 by Dawkins [23]. The good review on a current state of the theory of evolutionary games is given in [24,25,26].

- **Neural networks.** In 1982 Hopfield [27] offered model of the neural network based on pairwise interactions between neurons. He also found Lyapunov's function for this model. From the formal point of view this equations correspond to the generalized Lotka-Volterra model [16]. Since the work of J.J. Hopfield, neural networks have become one of the most important concepts in the sciences of interacting systems. For example one can see [28-31].

- **Biological networks.** In 2006 E.Schneidman with coauthors [32] showed that for the description of functioning retinal ganglion cells as they respond to natural movies has enough accounting of pairwise interactions. This work was based on use of Izing model and the principle of a maximum
entropy. The papers by Roudi et al. [33,34] and Tkacik et al. [35] provide a survey of recent activity in biological networks.

- **Financial markets dynamics.** Recent results of Bury [36,37] have shown that the market microstructure capturing almost all of the available information in the data of stock markets does not involve higher order than pairwise interactions. Farmer in [38] discusses a set of differential equations that describe the flow of capital and pointed out that these equations are equivalent to the generalized Lotka-Volterra equations with pairwise interactions.

  **Social interactions models.** Zhao et al. in [39] consider a general model of pairwise human social interaction consisting of N agents representing interacting in social network. Lou et al. in [40] analyze standard generalized additive models (GAMs) with pairwise interactions.

  As we have briefly discussed above, different types of pairwise interactions have played an important role in the analysis of many natural phenomenon.

### 2. The Theoretical Results.

#### 2.1. Equations of macrosystems with nonlinear pairwise interactions.

One of the important assumptions in classic Lotka-Volterra and replicator systems is that the functional responses are linear functional responses. Usually additive linear interactions models for predicting combined interactions effects cannot account for non-linearities in combined functional response introduced by non-trophic interactions Ayala et al [41]. Exponential and power law responses were used to model social dynamics by Beechman and Farnsworth [42]. The PhD Theses by Merrifield [43] provides in Chapter 5 a review on the concept of a force including nonlinear in different interacting systems.

Consider a macrosystem formed by a sufficiently large number $N$ of interacting objects. The classical definition of macrosystems is as follows: these are systems in which a chaotic behavior at the microlevel transforms into a deterministic behavior at the macrolevel. Suppose that, at a moment $t$, the macrosystem under consideration contains $n$ different types of objects and the number of objects of type $i$ is $x_i(t)$, where $i = 1,...,n$, and $\sum_{i=1}^{n} x_i(t) = N(t)$. Consider the relative numbers $p_i(t) = x_i(t)/N(t)$ of various objects types. Obviously, $\sum_{i=1}^{n} p_i(t) = 1$, i.e., $p(t) \in \sigma^n_p = \{ p \in \mathbb{R}^n : p_i \geq 0, i = 1,2,...,n, e^T p = 1 \}$, where $\sigma^n_p$ is the standard simplex in Euclidean $n$-space $\mathbb{R}^n$ and, $e$ is the vector of ones. Thus, the state of such a macrosystem at each moment $t$ is determined by the vector $p(t) = (p_1(t),...,p_n(t))$. 
We consider the systems in which only pairwise interactions occur during a short time interval \((t, t + \Delta t)\); in other words, those systems in which simultaneous interactions of more than two particles are impossible. Such a constraint is fairly natural and has found a recent confirmation in [6,32]. We make the following two assumptions about interactions between objects in a macrosystem [44].

**Hypothesis 1.** The interaction between objects of types \(i\) and \(j\) is characterized by the so-called interaction strength, which we denote by \(w_{ij}\) and regard as a quantitative characteristic of the effect of the interaction between two objects of types \(i\) and \(j\) on the rate of change of the relative number \(p_i(t)\) of objects of type \(i\).

**Remark 1.** Apparently, the notion of an interaction strength first appeared in mathematical ecology. A fairly detailed study of this notion and a survey of related results are contained in [45].

**Remark 2.** Note that the definition given above does not imply that \(w_{ij} = w_{ji}\). We also emphasize that the asymmetry of results of interaction plays an essential role in many cases [46].

**Hypothesis 2.** For each macrosystem under consideration, there exists a set of probability distribution functions \(f_i(p_i), \ i = 1, \ldots, n\), which determine the probability of the interaction of each object of type \(i\) with any other object in the macrosystems. Thus, the probability of the pairwise interaction between objects of types \(i\) and \(j\) is determined by the product \(f_i(p_i)f_j(p_j)\).

**Remark 3.** The probability distribution functions \(f_i\) (nonlinear response functions) obeys the conditions \(f_i(0) = 0, \ \partial f_i/\partial p_i > 0\) for \(p_i > 0\), \ \partial f_i/\partial p_i \geq 0\) for \(p_i = 0\) and \(f_i(1) = 1\).

It follows from the two hypothesis stated above that, for small \(\Delta t\), the dynamics of the system is determined by the relation

\[
p_i(t + \Delta t) = p_i(t) + \Delta t \sum_{j=1}^{n} w_{ij} f_i(p_i(t)) f_j(p_j(t)) + o(\Delta t),
\]

subject to the constraints \(\sum_{i=1}^{n} p_i(t) - 1 = 0\) and \(f_i(0) = 0\), for \(i = 1, 2, \ldots, n\) which ensure the invariance of the simplex \(\sigma^n_p\) and all of its faces. Using an approach common in analytic mechanics, we treat the system under consideration as a nonfree system subject to the ideal holonomic constraints

\[
f_i(p_i)\left(\sum_{i=1}^{n} p_i(t) - 1\right) = 0, \ \ i = 1, \ldots, n
\]

and apply the principle of replacing constraints by reaction forces to this system [47]. Passing to the limit as \(\Delta t \to 0\) in (1) and taking into account the obvious identity \(\sum p_i = 0\) for the sum of
derivatives, we obtain the following equation for the constraint multipliers (the indeterminate Lagrange multipliers) $\lambda_i$:

$$\sum_{ij} w_{ij} f_i(p_i) f_j(p_j) - \sum_{i=1}^n \lambda_i f_i(p_i) = 0$$

This equation has two obvious solutions, the trivial solution $\lambda_i = \sum_{j=1}^n w_{ij} f_j(p_j)$ and the nontrivial solution

$$\lambda_i = \frac{\sum_{j=1}^n w_{ij} f_i(p_i) f_j(p_j)}{\sum_{i=1}^n f_i(p_i)}, \quad i = 1, \ldots, n$$

Using the nontrivial solution and denoting the sum $\sum_{i=1}^n f_i(p_i)$ by $\theta(p)$, we obtain the following system of differential equations determining the evolution of the probability distribution $p(t)$:

$$\dot{p}_i = f_i(p_i) \left( \sum_{j=1}^n w_{ij} f_j(p_j) - \theta^{-1}(p) \sum_{j=1}^n w_{ij} f_i(p_i) f_j(p_j) \right), \quad i = 1, \ldots, n \quad (3)$$

This system of equations was first obtained from balance considerations in [48] without the use of the Lagrange method. In what follows, it is convenient to pass to the matrix form. System (3) is written in this form as

$$\dot{p} = D(f) \left( Wf - \theta^{-1}(p) \langle f, Wf \rangle \right) \quad (4)$$

Here, $f(p)$ is the vector $(f_1(p_1), \ldots, f_n(p_n))$; $D(f) = \text{diag}(f_1, f_2, \ldots, f_n)$; $W = (w_{ij})$ is the matrix of interactions; and $\theta(p) = \langle e, f(p) \rangle$, where $\langle \cdot, \cdot \rangle$ denotes inner product. Obviously, since $\langle \dot{p}(t), e \rangle = 0$ and $f_i(0) = 0$ it follows that the simplex $\sigma^+_p$ and each of its faces are invariant sets for system (4). Note that system (4), as well as the generalized Lotka-Volterra equations [15, 16], determine the dynamics of objects with nonlinear pairwise interactions; the matrix $W$ determines the structure of interactions, and the response functions determine their types. The elements of the matrix $W$ are generalized interaction strengths, i.e., by analogy with the thermodynamics of irreversible processes, are "reasons" causing changes in the speed of flows [49].

2.2 Escort distributions

Let us rewrite Eqs. (4) as

$$\dot{p} = \theta D(f) \left( Wf - \theta^{-1}(p) \langle f, Wf \rangle \right) \quad (5)$$

where $E(p) = \theta^{-2}(p) \langle f, Wf \rangle$. Using the terminology of neural networks theory [27], we refer to $E(p)$ as the energy function of the macrosystem under consideration. It is important to point out that in the population genetic theory this function named as population fitness [19], and in the evolutionary game
theory as utility function [22]. System (5) and the structure of energy function $E(p)$ naturally determine the introduction of new additional variables:

$$x_i(p) = f_i(p_i)^{0-1}(p), \quad i = 1, \ldots, n. \tag{6}$$

Obviously, $x = (x_1, x_2, \ldots, x_n) \in \sigma^n_x = \left\{ x \in \mathbb{R}^n : x_i \geq 0, e^T x = 1 \right\}$ for $p \in \sigma^n_p$. The indices $x$ and $p$ are used in the notation of simplices in order to avoid confusion. Consider change (6) in more detail. If this is a diffeo-morphism, then it can be regarded not only as a simplifying change of variables customary in the theory of differential equations but also as the definition of a set of quantities with particular physical meaning.

The somewhat different change of variables $x = Sp(\|Sp\|)^{-1}$, $p \in \sigma^n_p$, where $S \geq 0$ is a nonnegative nonsingular matrix and $\|\cdot\|$ is the Holder $r$-norm, was considered in [15] for the classical Fischer equations (i.e., in the case $f_i(p_i) = p_i$). Under such a change, the motion of the transformed system occurs on the segment of the hypersphere $\|x\| = 1$ whose boundaries are determined by the relation $p \in \sigma^n_p$. In particular, this transformation has made it possible to obtain an expression for the weighted Kullback-Leibler relative entropy [15,16].

However, in the case of nonlinear pairwise interactions, the approach based on "mixing" variables leads to no substantial progress. Let us return to (6). Apparently, for the case of power functions $f_i$ of the same degree, this change was first used in 1940 in research in quantum mechanics [50]. Subsequently, in analysis of the thermodynamics of chaotic systems [51], it was proposed to use the name "escort distributions" for such distributions, which quite adequately reflects the fact of the matter. A fairly comprehensive survey of escort distributions is contained in [52]. In [48], it was shown that the problem of the existence of a one-to-one transformation into (6) in the complete space $\mathbb{R}^n$ reduces to a generalized functional Cauchy equation and has a unique solution if and only if all functions $f_i(p_i)$ are power functions of the same degree. However, considering mapping (6) on its natural domain, we can prove that this is a $C^1$ diffeomorphism on this domain.

**Theorem 1.** [53] Under the conditions (remark 3) used for the response functions $f_i$, for $p \in \sigma^n_p$, there exists a one-to-one inverse mapping to (6), which is defined by

$$p_i = f_i^{-1}(x_i)/\sum_{j=1}^n f_j^{-1}(x_j), \quad i = 1, 2, \ldots, n, \tag{7}$$

where $f_i^{-1}(\cdot)$ denotes the function inverse to $f_i(\cdot)$.

**Remark 4.** Recall that the notation $f_i^{-1}(\cdot)$ is used for both functions inverse in the sense of function theory and functions inverse in the algebraic sense. It is always clear from the context what is meant.
Proof. Suppose given a distribution $p \in \sigma^*_p$ and the related mapping $F : \sigma^*_p \to \sigma^*_x$ defined by (6). We must prove that there exists a one-to-one inverse $F^{-1} : \sigma^n_x \to \sigma^n_p$ and it is defined by (7). Let us denote the unit $n$-cube by $\mathbb{K}^n_y = \{ y : 0 \leq y_i \leq 1, i = 1, ..., n \}$. By virtue of the assumptions made above, we have $y = f (p) \in \mathbb{K}^n_y$. Let us decompose the mapping $F$ into two stages as $F : \sigma^*_p \xrightarrow{f} \mathbb{K}^n_y \xrightarrow{g} \sigma^*_x$, where

$$g_i (y) = y_i / \sum_{j=1}^n y_j = x_i \quad i = 1, ..., n \quad (8)$$

The points $y$ determine a pencil of rays centered at $0 = (0, ..., 0)$ in $\mathbb{R}^n_+$, and relations (8) specify, in an obvious way, a system of homogeneous coordinates. Indeed, (8) implies $\frac{y_i}{x_i} = \frac{y_j}{x_i}$ for $i, j = 1, 2, ..., n$, i.e., $y_i = \alpha x_i$, where $\alpha \in [0, \infty)$, which determines a perspective correspondence between the rays in the pencil and the points of the simplex $\sigma^n_x$. Identifying points $y \in \mathbb{K}^n_y$ with the corresponding rays, we obtain a one-to-one correspondence between the points of the simplex $\sigma^n_x$ and segments of rays being elements of the cube $\mathbb{K}^n_y$, which leads to the diagram

$$\begin{array}{ccc}
\sigma^*_p & \xrightarrow{f} & \mathbb{K}^n_y \\
g^{-1} \uparrow & & \downarrow g \\
\mathbb{K}^n_z & \xleftarrow{f^{-1}} & \sigma^*_x
\end{array}$$

where $z = (z_1, ..., z_n)$, $z_i = f_i^{-1} (x_i)$. Since the functions $f_i$ monotonically increase on the interval $[0, 1]$, it follows that there exist single-valued inverse functions $f_i^{-1}$ defined and increasing on the same interval. The mapping $g : \mathbb{K}^n_y \to \sigma^n_x$ determining the perspective correspondence has a one-to-one inverse $g^{-1} : \mathbb{K}^n_z \to \sigma^n_p$ as well, which proves the commutativity of diagram and completes the proof of the theorem.

2.3. Equilibriums

To go further, we need the following assertion.
Statement 1. System (4) is invariant with respect to the replacement of the interaction matrix $W$ by a perturbed matrix $W_\zeta = (W + \zeta^T(p))$, where the components of the vector function $\zeta(p) = (\zeta_1(p), ..., \zeta_n(p)) : \sigma^n_p \rightarrow \mathbb{R}^n$ are bounded on $\sigma^n_p$.

This statement was proved in [15] for the case $\zeta(p) = \text{const}$. This proof carries over in an obvious way to the case of any bounded functions $\zeta_i(p) \quad i = 1, 2, ..., n$. The meaning of this statement is in part clarified by the following lemma proved in [15].

Lemma. Let $A$ be a linear operator on the complex space $\mathbb{C}^n$ having a characteristic number $\lambda_k(A)$ with simple elementary divisor and eigenvector $c$, and let $\zeta \in \mathbb{C}^n$ be an $n$-vector. Then the characteristic numbers of the perturbed operator $B = A + \langle \zeta, c \rangle$ are preserved up to the translation $\lambda_k(B) = \lambda_k(A) + \langle c, \zeta \rangle$ for any $\zeta \in \mathbb{C}^n$.

The assertion stated above is useful because the phase space of system (4) is the standard simplex $\sigma^n_p$ contained in the subspace of $\mathbb{R}^n$ orthogonal to the unit vector $e$.

In due time, this fact served as a foundation for the creation of information geometry, which is the science studying various Riemannian metrics on manifolds of probability distributions. Apparently, the first work in this direction was paper [54], in which a Riemannian metric on $\text{Int} \sigma^n_p$ based on the Fisher information metric was introduced. At present, this metric is known as the Fisher-Rao Riemannian metric on $T_p \sigma^n_p$ is given by

$$M_p (u, v) = \sum_{i=1}^n \frac{u_i v_i}{p_i}, \quad u, v \in T_p \sigma^n_p,$$

where $T_p \sigma^n_p$ is the tangent space in point $p$ to $\text{Int} \sigma^n_p$, i.e. $u^T e = v^T e = 0$. In fundamental monograph by Chentsov [55], it was proved, in particular, that this metric is a unique Riemannian metric invariant with respect to the Markov morphism. This metric usually used for classical Fisher equations. If follows in the standard way that if we deal with system (4) then a Riemannian metric is given by

$$M_f (u, v) = \sum_{i=1}^n \frac{u_i v_i}{f_j(p_i)}, \quad u, v \in T_p \sigma^n_p,$$

Results of further studies on the methods of information geometry of divergence functions are given in [56]. Statement 1 proved above makes it possible to investigate system (4) of probability distributions evolution equations with taking into account the fact that we deal with the restriction of system (4) to the simplex $\sigma^n_p$. 
We proceed to consider questions related to the existence of a nontrivial equilibrium point \( \hat{p} \in \text{Int} \sigma^n \) for system (4). In [15, 48], it was proved that \( \hat{p} > 0 \) exists if and only if all components of the vector \( W^{-1}e \) are of the same sign. It is clear already from this result that properties of the inverse interaction matrix play a substantial role in the evolution of macrosystems. Statement 1 implies that the coordinates of a nontrivial equilibrium point remain invariable \( \forall \zeta \in \mathbb{R}^n \). Let us prove this directly.

**Statement 2.** [53] If system (4) has a nontrivial equilibrium point \( \hat{p} \in \text{Int} \sigma^n \), then

\[
\frac{\hat{f}}{\langle \hat{f}, e \rangle} = \frac{W e^{-1}}{\langle e, W^{-1} e \rangle} = \frac{W_{\zeta^{-1}}^{-1}}{\langle e, W_{\zeta^{-1}}^{-1} e \rangle} = \hat{x} \quad \forall \zeta(p) \in \mathbb{R}^n
\]  

**Proof.** We use the following notation:

\[W^{-1}e = a, \quad \langle e, W^{-1}e \rangle = b,\]  
i.e. \( \hat{x} = \frac{a}{b} \). Using an expression for the matrix \( W_{\zeta^{-1}}^{-1} \) given in [57] (p. 139), we obtain

\[
(W + e\zeta^T)^{-1} = W^{-1} - \frac{a\zeta^T}{1 + \zeta^T a} W^{-1}
\]

therefore, \( W_{\zeta^{-1}}^{-1}e = \frac{a}{1 + \zeta^T a} \). Multiplying this relation by \( e^T \) on the left, we obtain

\[
e^T W_{\zeta^{-1}}^{-1}e = \frac{e^T a}{1 + \zeta^T a} = \frac{b}{1 + \zeta^T a},
\]

i.e., \( \frac{W_{\zeta^{-1}}^{-1}e}{e^T W_{\zeta^{-1}}^{-1}e} = \frac{a}{b} = \hat{x} \), \( \forall \zeta(p) \in \mathbb{R}^n \), which completes the proof of the statement.

**Corollary 1.** The escort distribution equilibrium value \( \hat{x} \) depends only on the properties of the inverse interaction matrix, and the equilibrium vector \( \hat{p} \) is determined from the systems (7) and (9):

\[
\hat{p}_i = f_i^{-1}(\hat{x}_i) / \sum_{j=1}^n f_j^{-1}(\hat{x}_j), \quad i = 1, \ldots, n
\]  

The distribution (10) is sometimes called the Gibbs distribution. The normalization constant is determined as usual by

\[
Z = \sum_{j=1}^n f_j^{-1}(\hat{x}_j)
\]

and is conventionally called the partition function.

2.4. Energy-like Lyapunov-Meyer functions
Following the stability postulate, we state the following theorem, obtained in [53]

**Theorem 2.** If there exists a vector \( \zeta \), for which \( W_\zeta = W_\zeta^T \), then the energy function of the system

\[
E_\zeta(p) = \langle f(p), W_\zeta f(p) \rangle \theta^{-2}(p)
\]

(12)

is its Lyapunov-Meyer function.

The proof of this theorem is based on the fact that, under the assumptions of the theorem, system (5) can be written in the gradient form

\[
\dot{p} = D(f(p)) \nabla E_\zeta(p),
\]

where \( \nabla E_\zeta(p) = (\partial E_\zeta(p)/\partial p_1, \ldots, \partial E_\zeta(p)/\partial p_n) \) is the gradient operator and \( D(f(p)) \) is a diagonal matrix with positive elements. We point out the obvious fact that the energy function contains no information about the equilibrium of the system (5), and the landscape determined by this function makes it possible to perform a complete qualitative study of the system trajectories behavior in its phase space, \( \sigma^n_p \), including its faces. We emphasize that this result is a substantial generalization of the fundamental theorem of natural selection [19].

**Remark 5.** The symmetry conditions on the interaction matrix in Theorem 2 are related in an obvious way to the well known Onsager reciprocity relations [49].

Theorem 2 has the following useful corollary.

**Corollary 2.** If the interaction matrix \( W \) is tridiagonal, then there exists a vector \( \zeta \in \mathbb{R}^n \) such that the energy function \( E_\zeta \) of the system is a Lyapunov-Meyer function.

**Proof.** The system of equations \( W + e\zeta^T = W^T + \zeta e^T \) in the case of a tridiagonal matrix \( W \) gives \( (n-1) \) linear equations determining the components of the vector \( \zeta \), which leaves some freedom in the choice of the required symmetrizing vector \( \zeta \).

Thus, if pairwise interaction occurs only between “nearest neighbors,” then the evolution of the macrosystem is of gradient type.

Also from (9) and (12) we can receive the expression for equilibrium energy

\[
\hat{E} = E(\hat{p}) = \langle e, W^{-1}e \rangle^{-1}
\]

(13)

Note, that \( \hat{E} \) depends only from interaction matrix.

### 2.5. Entropy-like Lyapunov-Meyer functions

**Theorem 3.** [53] If system (4) has a nontrivial equilibrium point \( \hat{p} \in \text{Int}\sigma^n_{\hat{p}} \) and the matrix 

\[
(W_\zeta^T + W_\zeta), \quad \forall \zeta \in \mathbb{R}^n
\]

has \( n-1 \) negative characteristic numbers, then the function
where \( C \) is a constant, is a Lyapunov-Meyer function for system (4) on \( \text{Int} \Sigma_p^n \), and the energy function \( E(p) \) of the system attains its maximum value \( E(\hat{p}) \) as \( t \to \infty \).

**Proof.** For the derivative of the function \( H(p) \) along the trajectories of system (4), we have

\[
\dot{H}(p) = \dot{f}^T (Wf - e\theta^{-1} \langle f, Wf \rangle)
\]  

Let us rewrite (15) in terms of the escort distribution \( x(t) \). Taking into account the relation \( x^T e = 1 \), we obtain

\[
\dot{H}(p) = \hat{\theta} \langle \hat{x}^T Wx - x^T Wx \rangle
\]  

Thus, we must prove that the hyperplane \( \hat{E}(x) = \hat{x}^T Wx \) is tangent to the surface \( E(x) = x^T Wx \) at the point \( x = \hat{x} \). Let us introduce the auxiliary variables \( V_i = x_i - \hat{x}_i \), \( i = 1, \ldots, n \). The vector \( v \) belongs to the "displaced" simplex \( \sigma_v^n = \left\{ v : e^T v = 0, \hat{x}_i < v_i < 1 - \hat{x}_i \right\} \), which is the tangent space to \( \text{Int} \Sigma_v^n \). After simple transformations, we obtain

\[
\dot{H}(p) = -\hat{\theta} \langle v, Wv \rangle.
\]  

According to the conditions of the theorem and Kingman's result [58], the quadratic form \( \langle v, Wv \rangle \) is negative definite on \( \Sigma_v^n \), i.e., \( \dot{H}(p) > 0 \) for \( p \neq \hat{p} \), which proves the first part of the theorem. Relation (15) with (5) taken into account can be rewritten as

\[
\dot{H}(p) = \hat{\theta} \langle \hat{E}(\hat{p}) - E(p) \rangle \geq 0,
\]  

where \( \hat{E}(\hat{p}) = \langle f, Wf \rangle \hat{\theta}^{-1} \theta^{-1} \), which completes the proof of the theorem.

**Remark 6.** As is easy to see, relation (18) determines the entropy production rate in the system. In [46], it was shown that if the interaction matrix \( W \) satisfies the condition \( W^T W^{-1} e = e \), then \( \hat{E}(\hat{p}) = E(\hat{p}) \), which sometimes simplifies the consideration of relation (18).

**Remark 7.** It is obvious that entropy function (14) has the decomposable form in the sense of [2].

Consider properties of the function \( H(p) \). We begin with a quotation from [55]: "The most important information invariant of a pair of probability measures is relative entropy. ... This characteristic is asymmetric; therefore, we call it the information deviation of one distribution from the other... ." Chentsov's book [55] was written in 1972. Since then, many terminological changes have occurred in this area; they are summarized in the recently published Dictionary of Distances [59]. We shall use the term relative entropy, or the function of information divergence.
The function $H(p)$ with $p \in \sigma^n_p$ attains its maximum value at the point $p = \hat{p}$, where $H(\hat{p}) = C$; i.e., $H(p) < C \quad \forall p \neq \hat{p}$, the covector $H(p)$ is $\nabla H(p) = \left(\frac{\partial f_i}{\partial p_i} (p_i)\right)$, and the Hessian is $\mathcal{H}(H(p)) = -\text{diag} \left( \frac{\partial f_i}{\partial p_i} \right)^2$. Since $\frac{\partial f_i}{\partial p_i} > 0$, it follows that the Hessian is negative definite, and the function $H(p)$ is concave on $\sigma^n_p$. Thus, $-H(p)$ can be regarded as relative entropy, or the function of information divergence between the distributions $p$ and $\hat{p}$.

Note that the relative entropy $H(p)$ introduced by us depends on the interaction matrix only via the equilibrium point $\hat{p}$ of the system and is largely determined by the form of the response functions. This allows us to use relation (14) for solving the "inverse" problem of finding the response functions corresponding to a given entropy function $H(p)$.

In the results section we use Theorem 3 to construct entropy functions and “distances” between probability distributions. First, we note that, if $G : \mathbb{R}^1 \to \mathbb{R}^1$ is a monotonically increasing smooth function and $C_1$, $C_2$ and $C_3$ are the numbers, then the function

$$H_G = C_1 G(C_2 H(p)) + C_3$$

is also a Lyapunov-Meyer function for system (4). Clearly, the domain of $G$ coincides with the range of $\left(H(p) + C\right), p \in \text{Int}\sigma^n_p$.

### 2.6. Distance measures

Following Shannon information measure idea let us introduce $S(p)$ as negative entropy defined by:

$$S(p) = -H(p) = -\sum_{i=1}^{n} \int_{\hat{p}_i}^{p_i} \frac{\hat{f}_i(x)}{f_i(x)} dx, \quad p \in \text{Int}\sigma^n_p$$

(20)

It is obvious by the previous statement that the function $S(p)$ is defined on the $\text{Int}\sigma^n_p$, and we have to use directional derivative of $S$ for its analysis. However in this case it is easy to use the following evident identity

$$\lambda \sum_{i=1}^{n} \int_{\hat{p}_i}^{p_i} \frac{dx}{f_i(x)} = \lambda \sum_{i=1}^{n} \left(p_i - \hat{p}_i\right) \equiv 0 \quad \forall p, \hat{p} \in \text{Int}\sigma^n_p$$

and using (20) we clearly have, the following relationship
\[ S(p) = -\sum_{i=1}^{n} \int_{f_i}^{\hat{f}_i} \left( f_i(x) + \lambda \right) dx \quad \forall \lambda \in \mathbb{R}^1 \]

Since, by the evident equality \( \frac{\partial S(p)}{\partial p_i} = 0 \), on \( p = \hat{p} \), we have that \( \lambda = -1 \). It is clear, that \( \lambda \) is a type of Lagrange multiplier. Therefore:

\[ S(p) = \sum_{i=1}^{n} \int_{f_i}^{\hat{f}_i} \left( 1 - \frac{\hat{f}_i}{f_i(x)} \right) dx \quad p \in \text{Int} \sigma_{p}^{n} \quad (21) \]

The gradient of \( S(p) \) on \( p \in \text{Int} \sigma_{p}^{n} \) is given by vector of partial derivatives

\[ \nabla S(p) = \left( 1 - \frac{\hat{f}_i}{f_i(p_i)} \right) \quad (22) \]

with strictly monotonically increasing each component. Hessian of \( S(p) \) on \( \text{Int} \sigma_{p}^{n} \) is

\[ \mathcal{H}(p) = \text{diag} \left( \hat{f}_i \frac{\partial f_i}{\partial p_i} f_i^{-2}(p_i) \right) \]

Since \( \frac{\partial f_i}{\partial p_i} > 0 \) it follows that the Hessian is positive definite, and the function \( S(p) \) according to definition is convex on \( \text{Int} \sigma_{p}^{n} \). As observed by Ball and Chen [60] entropy and convexity have played an important role in many areas of mathematics. Continuing this line of reasoning, we see that \( S(p) \) is an distance measure between probability distributions \( p \) and \( \hat{p} \).

Another very well known definition of convexity Jensen [61], Hardy et al. [62] is the next Jensen inequality

\[ S(p) - S(q) - \langle \nabla_q S(q), p-q \rangle \geq 0 \quad p, q \in \text{Int} \sigma_{p}^{n} \quad (23) \]

It is obvious that the expression from left-side inequality is so called Bregman divergences usually denote by \( B_S(p,q) \). This name was first given the by Censor and Lent [63]. Bregman divergence or Bregman distance [64] is similar to a metric, but does not satisfy the triangle inequality nor symmetry. Using inequality (23) we can receive new weighted distance measure between two probability distributions \( p \) and \( q \).

\[ B_S(p,q) = \sum_{i=1}^{n} \left( \int_{f_i}^{q_i} \frac{\hat{f}_i}{f_i(x)} dx + \frac{\hat{f}_i}{f_i(q_i)} (p_i - q_i) \right) \geq 0 \]

Evidently that \( B_S(p,q) = 0 \) for \( p = q \), and \( B_S(p,\hat{p}) = S(p) \).

Another measures of information was proposed by Renyi in 1960 [65]. Today this measures is known as f-divergences. In 1963 Csiszar [66] and Morimoto [67] discovered this functions independently and now it is more known as Csiszar divergences.
\[ D_f(p,q) = \sum_{i=1}^{n} q_i h\left( \frac{p_i}{q_i} \right) \]

where \( h \) is a convex function on \((0,\infty)\) such that \( h(1) = 0 \)

Note that this function have found numerous application during the past years, and many works are devoted to an analysis of \( D_f(p,q) \) [68]. An analysis of relations between the functions (14) and (21) introduced by us and f-divergences will be the object of next paper.

2.7. Legendre-Donkin-Fenchel transformations

The Legendre-Donkin-Fenchel (LDF) transformation is a mathematical concept of great significance to thermodynamics, mechanics and probability theory is named after Legendre [69] Donkin [70] and Fenchel [71]. LDF-transformation is a way for representing a function in terms of its first derivative.

Let us denote \( l_i = \frac{\partial S(p)}{\partial p_i} \). Since \( l_i \) is strictly monotonically increasing for \( p_i \), this makes it clear that \( l_i \) and \( p_i \) are mutually reciprocal. The transformation from vector \( p \) to vector \( l \) is called LDF-transformation. We can find a convex function \( L \) of \( y \) defined by

\[ L(l) = \sum p_i l_i - S(p) \] (24)

when \( l \) and \( p \) are respective coordinates of the same point, and the inverse transformation from \( l \) to \( p \) is given by the gradient

\[ p = \nabla L(l) \] (25)

Concerning (22) we clearly have

\[ l_i = \frac{\partial S(p)}{\partial p_i} = 1 - \frac{\hat{f}_i}{f_i(p_i)} \]

and

\[ p_i = f_i^{-1}\left( \frac{\hat{f}_i}{1-l_i} \right) = \frac{\partial L}{\partial l_i} \]

for \( l_i \in (-\infty, 1 - \hat{f}_i) \). We also denote \( \partial L/\partial l_i = \phi(l_i) \) and after simple counting we have the following relationship

\[ L(l) = \sum_{i=1}^{n} l_i \phi(l_i) + \int_{0}^{\frac{\hat{f}_i}{f_i}} \left( \frac{\hat{f}_i}{f_i} - 1 \right) dx \] (26)

After simple computation one can check that \( \partial L/\partial l_i = p_i \). Another very well known form of LDF-transformation, together with (24) is
Follow Amari [68,72] and definition (24) one can define a divergence function (distance measure) between two points \( p \) and \( q \) on \( \mathbb{S}_p^n \) in the dual coordinates. This divergence function will be similar to the squared Riemannian distance. Additional new recent information on LDF-transformation is presented in Nielsen [73] and Zia et al [74] papers.

3. Examples

3.1. Boltzmann entropy

Consider the Boltzmann-Shannon entropy \( H(p) = -\sum_{i=1}^{n} p_i \ln p_i \). Taking into account the background of the system (4) it is clear that the interaction matrix \( W \) is stochastic, i.e. \( We = e \) and the uniform distribution \( \hat{p} = e \) must be a nontrivial equilibrium point of the system, and all response functions \( f_i(\cdot) \) are the same. From the equation

\[
-\sum_{i=1}^{n} p_i \ln p_i = \sum_{i=1}^{n} \frac{p_i}{n} \int f_i(x) \, dx + C
\]

using the notation of Theorem 1 for nonlinear response functions, we obtain

\[
y_i = f_i(p_i) = (1 - \ln p_i)^{-1}, \quad i = 1, \ldots, n.
\]

Obviously that on the interval \([0, 1]\), functions (28) satisfy the conditions specified in Remark 3 and this is convex functions. It can be shown in the usual way that the inverse functions

\[
f_i^{-1}(y_i) = \exp(1 - 1/y_i), \quad y \in \mathbb{R}^n_y, \quad i = 1, \ldots, n
\]

It is clear, that (29) is concave functions. If we consider equations of evolution (4) with an arbitrary interactions matrix \( W \) satisfying the condition of Theorem 3, then instead of Jaynes [75] maximum-entropy principle we can receive the equilibrium probability distribution from (9) and (10):

\[
\hat{p}_i = \exp\left(-\frac{1}{\hat{x}_i}\right)/Z
\]

Recall that \( \hat{x}_i = \hat{E}W^{-1}e \), so \( \hat{x} \) is independent from the type of interactions.

Of most interest in this example are, certainly, functions (28) and (29), which make it possible to not only apply the developed theory of pairwise interaction system but also consider the problem of finding a possible equivalence relation between the Boltzmann kinetic equations based on pairwise interactions and system (4).
3.2. Power-law entropies

Suppose that the matrix $W$ satisfies the conditions of Theorem 3. As the functions $f_i(p_i)$, we consider power-law function

$$f_i(p_i) = p_i^{1-\alpha}, \quad \alpha < 1, \quad i = 1, \ldots, n$$  \hspace{1cm} (30)

Performing calculation we obtain from (14)

$$H(p) = \sum_{i=1}^{n} \left( p_i^\alpha \frac{\hat{p}_i^{1-\alpha}}{\alpha} - \frac{\hat{p}_i}{\alpha} \right) + C$$  \hspace{1cm} (31)

If $\hat{p} = e_\mu^{-1}$, then setting $C = \frac{1}{\alpha}$, we received power-law entropy function

$$H(p) = \frac{n^{(\alpha-1)}}{\alpha} \sum_{i=1}^{n} p_i^\alpha$$  \hspace{1cm} (32)

Next we apply (19) and show how we can receive, for example, Renyi and Tsallis entropy functions.

**Renyi entropy function.** Setting in (19) $C_2 = \alpha/n^{(\alpha-1)}$, $C_1 = (1-\alpha)^{-1}$, $G(\cdot) = \ln(\cdot)$ we obtain the Renyi entropy function [65]:

$$H_R = (1-\alpha)^{-1} \ln \left( \sum_{i=1}^{n} p_i^\alpha \right)$$  \hspace{1cm} (33)

**Tsallis entropy function.**

The reader will have no difficulty in showing the same way that Tsallis entropy function [76]:

$$H_T = \left( \sum_{i=1}^{n} p_i^\alpha - 1 \right)^{1/(1-\alpha)}$$  \hspace{1cm} (34)

Recall that function (34) was introduced by Tsallis for the purpose of constructing the thermodynamics of nonextensive systems, and the parameter $\alpha$ is a nonextensivity measure of the system. The creation of such a thermodynamics is necessary because large systems with so called long-range interactions cannot be described by the Boltzmann thermodynamics.

Note that expression (32) is valid as LMF for (4) only if $0 < \alpha < 1$. The lower bound is determined by the requirement that the improper integrals in (14) must converge and the upper bound follows from the equalities $f_i(0) = 0$.

3.3. Kullback-Leibler distance measure
In the following example, we consider one of the best known measures of distance between probability distributions. Suppose that the matrix $W$ satisfies the conditions of Theorem 3. We set $$f_i(p_i) = p_i$$ for $i = 1, 2, ..., n$. Then (14) gives the Kullback-Leibler relative entropy [77]

$$H_{\beta}(p) = \sum_{i=1}^{n} \hat{p}_i \ln \left( \frac{p_i}{\hat{p}_i} \right)$$

First time, this function was used as Lyapunov function to analyze the stability of equilibrium points of replicator systems in [78]. Note that, since the functions $f_i$ are linear in this case, we deal with the classical system of replicator equations. A detailed analysis of such equations was made in [15]; in particular, it was shown that, for such systems, we can remove the constraint $W = W^T$ and obtain results similar to those obtained above for matrices of the type $W = D_1BD_2$, where $D_1$ and $D_2$ are diagonal matrices, $D_1 D_2 > 0$, and $B = B^T$, i.e., for the class of so-called diagonally symmetrizable matrices. A necessary and sufficient condition for a real matrix $W$ to be diagonally symmetrizable was obtained in [16]. It is so far unclear whether the constraints $W = W^T$ can be removed in the case of nonlinear functions $f_i$.

4. Discussion

It is seen from the examples given above that many (and practically all) known entropy functions and distance measures may be obtain from entropy-like Lyapunov functions (14) and (19). We also emphasize that there exists a relation between the derivative of the function $H(p)$, which can be interpreted as a generalized entropy, and the function $E(p)$, which is often considered as an analog of the neural networks energy or population fitness. This relationship for entropy production was established by Pykh [44,53] for corresponding types of interactions matrix and has the next form:

$$\dot{H}(p) = \theta \left( \dot{E}(p) - E(p) \right) \geq 0$$

Note that it was Ilya Prigogine who first pointed out [5] the importance of the relationship between Lyapunov functions and entropy.

The approach developed in this paper makes it possible to associate each entropy functions with some set of dynamical systems which explicitly depend on the character of the interactions between the objects under examination, which, in the author's opinion, may be essential for solving particular problems.

Note that obtained results can be used in two ways.

1. The response functions can be found for previously known entropy characteristics. Relevant examples for the Shannon and Tsallis entropy were given in [79]. It is easy to show that this approach
yields response functions for all the entropy characteristics proposed in [80].

2. New entropy characteristic can be derived from some functions obeying the conditions stated for response functions. As an example, we consider the logistic function, which is widely applied in mathematical ecology and economics [46].

In conclusion, we emphasize that system (4) is based on the acceptance of only two hypotheses about the structure and the type of nonlinear pairwise interactions. However, results of the mathematical study of this system make it possible to obtain not only the classical Boltzmann-Shannon entropy but also any other known parametric entropy function.

We also mention that all results stated above were obtained by formally analyzing systems of generalized replicator equations, which arise in very diverse fields of natural sciences and, therefore, can serve as a basis for finding analogies between these domains of natural sciences.

Let us point out also that construction of LMF for general nonlinear systems is of great theoretical and practical interests [81]. If this function exists, that means that the system under considerations is a gradient system [82].

The reader may consider this paper as a step in the development of nonlinear pairwise interactions dynamical systems theory.

**Conflicts of Interest**

The author declare no conflict of interest.
REFERENCES

1. Chetaev, N.G. Stability and classical Physical Law. Proceedings of Kazan Aviation Institute. 1936, 6, 3-5 (in Russian).

2. Coleman, B.D.; Mizel, V.J. Existence of Entropy as a Consequence of Anymntotic Stability. Arch. Rational Mech. Anal. 1967, 25, 243-270.

3. Meyer, K. Energy Functions for Morse-Smale systems. Amer. J. Math. 1968, 90, 1031-1040.

4. Smale, S. On Gradient dynamical systems. Annals of Mathematics. 1961, 74, 199-206.

5. Prigogine, I. Time, structure and fluctuations. Nobel lecture, 8 december, 1977.

6. Bialek, W.; Rangqnatham, R. Rediscovering the power of pairwise interactions. 2007. arXiv:0712.4397 v1. [q-bio.QM].

7. Yamakawa, T.; Kreinovich, V. Why Fundamental Physical Equations Are of Second Order. – International Journal of Theoretical Physics. 1999, 38, 1763-1769.

8. Kreinovich, V. Fundamental properties of pair-wise interactions naturally lead to quarks and quark confinement: a theorem motivated by neural universal approximation results. UTEP-CS-99-41, 2000, 8 p.

9. Nijkamp, P. Gravity and Entropy Models: The State of the Art, New Developments. In Modeling Travel Demand and Urban Systems, Jansen, G.R.M., P.H.L. Bovy, J.P.J.M. Van Est and F. LeClerq, Eds., Gower, Aldershot, 1979; pp. 281-316

10. Hill, T.L. Thermodynamics of Small Systems; Benjamin, New York, 1963.

11. Carrete, J.; Valera, L.M.; Gallego L.J. Nonequilibrium nanothermodynamics. Physical review. E. 2008. 77, 022102.

12. Lotka, A.J. Analytical Note on Certain Rhythmic Relations in Organic Systems, Proc. Natl. Acad. Sci. U.S., 1920, 6, 410-415.

13. Volterra, V. Variazioni e fluttuazioni del numero d’individui in specie animali conviventi. Mem. Acad. Lineci Roma, 1926, 2, 31-113.

14. Wilson, A. Boltzmann, Lotka and Volterra and spatial structural evolution: an integrated methodology for some dynamical systems. J. R. Soc. Interface, 2008, 5, 865-871.

15. Pykh, Yu.A. Equilibrium and stability in the models of population dynamics; Nauka: Moscow, 1983. (in Russian)

16. Pykh, Yu.A. Lyapunov functions for Lotka-Volterra systems: an overview and problems. Proceedings of 5th IFAC Symposium "Nonlinear Control Systems", 2001; pp. 1655-1660.

17. Ising, E. Beitrag zur Theorie des Ferro-und Paramagnetismus, Hamburg dissertation, 1924.
18. Broderich, T.; Dudik, M.; Tkacik, G.; Schapire, R.; Bialek, W. Faster solutions of the inverse pairwise Ising problem. 2007. arXiv: 0712.2437v2[q-bio.QM].

19. Fisher, R. A. The genetical theory of nature selection; Clarendon Press: Oxford, 1930.

20. Pykh, Yu.A. On the stability in the models of population genetics. Dissertation paper, Leningrad state University, 1972, 17 p. (in Russian)

21. Eigen, M. Selforganization of Matter and the Evolution of Biological Macromolecules. Die naturwissenschaften, 1971, 58, 465-528.

22. Taylor, P.D.; Jonker L.B. Evolutionarily Stable Strategies and Game Dynamics. Mathematical biosciences, 1978, 40, 145-156.

23. Dawkins, R. The Selfish Gene; Oxford University Press: Oxford, 1976.

24. Hofbauer, J.; Sigmund, K. Bull. Evolutionary game dynamics. Am. Math. Soc. 2003, 40, 479-519.

25. Dyer, M.; Mohanara, V. Pairwise interaction games. In Lecture Notes in Computer Science, 2011, 6755, 159-170.

26. Cressman, R.; Tao, Y. The replicator equation and other game dynamics. PNAS, 2014, 111, 10810-10817.

27. Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA, 1982, 79, 2554-2558.

28. Fitzgerald, J.D.; Sharpee, T.O. Maximally informative pairwise interactions in networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 2009, September: 80(3Pt1):031914.

29. Uykan, Z. Fast-Convergent Double-Sigmoid Hopfield Neural Network as Applied Optimization Problems. IEEE Transactions on Neural Networks and Learning Systems, 2013, 24, 990-996.

30. Bereng, D.; Dolev, S.; Hanemann, A. Graph Degree Sequence Solely Determines the Expected Hopfield Network Pattern Stability. Neural Computation. 2015, 27, 202-210.

31. Klau, G.W. A new graph-based method for pairwise global network alignment. BMC Bioinformatics. 2009, 10, 9 p.

32. Schneidman, E.; Berry, I.M.; Segev, R.; Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. Nature. 2006, 440, 1007-1012.

33. Roudi, Y.; Aurell, E.; Hertz J.A. Statistical physics of pairwise probability models. Frontiers in Computational Neuroscience 2009, 3, article 22.

34. Roudi, Y.; Tyrcha, J.; Hertz, J. The Ising model for neural data: model quality and approximate methods for extracting functional connectivity. Phys. Rev. E. 2009, 79, 051915.

35. Tkacik, G.; Marre, O.; Amodei, D.; Scheidman, E.; Bialek, W.; Berry, M.J. Searching for Collective Behavior in a Large Network of Sensory Neurons. PLOS Computational Biology. 2014, 10, e1003408.
36. Bury, T. Market structure explained by pairwise interactions. 2014. arXiv:1210.8380 v4[q-fin.ST].

37. Bury, T. Statistical pairwise interaction model of stock market. 2014. arXiv:1206.4420 v6[q-fin.ST].

38. Farmer, J.D. Market force, ecology and evolution. Industrial and Corporate Change, 2002, 11, 895-953.

39. Zhao, K.; Karsai, M.; Bianconi G. Entropy of Dynamical social networks. PLoS ONE, 2011, 6, issue 12, e28116.

40. Lou, Y.; Caruana, R.; Gehrke, J.; Hooker, G. Accurate Intelligible models with pairwise interactions. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA, 2013; pp. 623-631.

41. Ayala, F. J.; Gilpin, M. E.; Ehrenfeld, J.G. Competition between species: theoretical models and experimental tests. Theor.Pop.Biol., 1973, 4, 331-356.

42. Beecham, J.A.; Famsworth, K.D. Animal Group Forces Resulting from Predator Avoidance and Competition Minimization. Journal of Theoretical Biology. 1999, 198, 533-548.

43. Merrifield, A.J. An Investigation of Mathematical Models for Animal Group Movement, Using Classical and Statistical Approaches., PhD Doctorate, University of Sydney, 2006.

44. Pykh, Yu.A. Entropy characteristics of macrosystems with nonlinear pair interactions, Doklady Mathematics, 2011, 84, 669-671.

45. Novak, M.; Wootton, T.J. Estimating nonlinear interaction strengths: an observation-based method for species-rich food webs. Ecology. 2008, 89, 2083-2089.

46. Pykh, Yu.A. Construction of entropy characteristics based on replicator equations with nonsymmetric interaction matrices, Doklady Mathematics, 2005, 72, 780-783.

47. Loitsyanskii, L.G.; Lur’e, A.I. Course in Theoretic Mechanics. Nauka: Moscow, 1983. (in Russia)

48. Pykh, Yu.A. Energy Lyapunov function for generalized replicator equations. Proceedings of International Conference “Physics and Control”, 2003, pp. 271-276.

49. Prigogine, I.; Stengers, I. Order out of Chaos. Man’s New Dialogue with Nature. N.Y., Bantam Books, 1984.

50. Husimi, K. Some Formal Properties of the Density Matrix, Proc. Phys. Math. Soc. Jpn., 1940, 22, 264-314.

51. Beck, C.; Schloegl, F. Thermodynamics of Chaotic Systems: An Introduction. 1993, Cambridge University Press (England).

52. Bercher, J.F. A simple probabilistic construction yielding generalized entropies and divergences, escort distributions and q-Gaussians. Physica A. Stat. Theor. Phys. 2012, 391, 4460-4469.
53. Pykh, Yu.A. Escort distribution, energy and relative entropy of macrosystems with nonlinear pair interactions, *Doklady Mathematics*, 2014, 90, 710-714.

54. Rao, C.R. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Math. Soc.* 1945, 37, 81-91.

55. Chentsov, N. *Statistical decision rules and optimal inference*, American Mathematical Society, R.I., 1982.

56. Amari, S.; Cichocki, A. Information geometry of divergence functions. *Bulletin of the Polish Academy of Sciences*. 2010, 58, 183-195.

57. Ikramov, Kh.D. *Problem Book on Linear Algebra*. Nauka: Moscow, 1975. (in Russian)

58. Kingman, J.F.C. A mathematical problem in population genetics. 1961, *Proc. Cambr. Phil. Soc.* 57, 574-582.

59. Deza, E.; Deza, M.M. *Dictionary of Distances*, Elsevier, Amsterdam, 2006.

60. Ball, J.M.; Chen, G.G. Entropy and Convexity for nonlinear partial differential equations. *Phil. Trans. R. Soc. A*. 2013, 371, 20120340.

61. Jensen, G.L.W.V. Sur les functions convexes et les ineqalites entre les valenrs moyennes. *Acta Matematica*. 1906, 30(1), 175-193.

62. Hardy, G.H.; Littlewood, J.E.; Polya, G. *Inequalities*; Cambridge, At the University Press. 1934.

63. Censor, Y.; Lent, A. An iterative row-action method for interval convex programming. *Journal of Optimization Theory and Applications*. 1981, 34, 321-353.

64. Bregman, L.M. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *Comp. Math. and Math. Physics*. 1967, 7, 200-217.

65. Renyi, A. On measures of entropy and information. In: Proc. Fourth Berkeley Symp. on Mathematics, Statistics and Probability, 1960, Vol.1, Univ. of California Press, Berkly-Los Angeles, 1961, pp.547-561.

66. Csizsar, I. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar. Tud. Akad. Mat. Kutato int. Kozl*. 1963, 8, 85-108. (in German)

67. Morimoto, T. Markov processes and the $H$-theorem. *J. Phys. Soc. Jap.* 1963, 12, 328-331.

68. Amari, S. Differential geometry derived from divergence functions: information geometry approach. In M.Deza, M.Petitjean, K.Markov, Eds.; *Mathematics of Distances and Applications*; ITHEA, Sofia, Bulgaria, 2012, 9-23.

69. Legendre, A.M. L’intégration de quelques equations aux differences Partielles. *Memoires de l’Academie Royale des Sciences*, 1787, 309-351.
70. Donkin, W.F. On a class of differential equations, including those which occur in dynamical problems. *Philosophical Transactions of the Royal Society of London*. 1854, 144, 74-113.

71. Fenchel, W. On conjugate convex functions. *Canad. J.Math*. 1949, 1, 73-77.

72. Amari, S. Information geometry and its applications: convex function and dually flat manifold. In F.Niclsen (ed.), *ETVC 2008*, LNCS 5416; Springer-Verlag Berlin Heidelberg, 2009, 75-102.

73. Nielsen, F. Legendre transformation and information geometry. CIG-MEMO2, 2010.

74. Zia, R.K.; Redish, E.F.; McKay, S.R. Making sense of the Legendre transform. *Am. J. Phys.*, 2009, 77, 614-625.

75. Jaynes, E.T. Information Theory and Statistical Mechanics. *Physical Review*. 1957, 106, 620-630.

76. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys*. 1988, 52, 479-487.

77. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. of Math. Stat*. 1951, 22, 79-87.

78. Pykh Yu.A. On stability of a class of systems of nonlinear differential equations. *Soviet. Math. Dokl*. 1972, 13, 470-472.

79. Pykh Yu.A. Construction of entropy characteristics based on Lyapunov energy functions. *Doclady Mathematics*. 2004, 69, 355-358.

80. Esteban, M.D.; Morales, D. A summary of entropy statistics. *Kybernetica*, 1995, 31, 337-346.

81. Yuan, R.; Ma, Y.; Yuan, B.; Ao, P. Constructive proof of global Lyapunov function as potential function. 2010. arXiv:1012.2721 v1[nlin.CD].

82. Barta, T.; Chill, R.; Fasangova, E. Every ordinary differential equation with a strict Lyapunov function is a gradient system. *Monatshefte für Mathematik*. 2012, 166, 57-72.