Notions of the ergodic hierarchy for curved statistical manifolds

Ignacio S. Gomez

IFLP, UNLP, CONICET, Facultad de Ciencias Exactas, Calle 115 y 49, 1900 La Plata, Argentina

H I G H L I G H T S

- IGEH quantifies the statistical independence between the microvariables.
- Multivariate Gaussian distributions and quadratic Hamiltonians are one-to-one.
- IGEH is compatible with the criterium of global chaos given by the scalar curvature.
- Phase transitions in statistical models can be characterized by means of IGEH levels.
- Controlling phase transitions by external parameters of the covariance matrix.

A R T I C L E   I N F O

A B S T R A C T

We present an extension of the ergodic, mixing, and Bernoulli levels of the ergodic hierarchy for statistical models on curved manifolds, making use of elements of the information geometry. This extension focuses on the notion of statistical independence between the microscopical variables of the system. Moreover, we establish an intimately relationship between statistical models and families of probability distributions belonging to the canonical ensemble, which for the case of the quadratic Hamiltonian systems provides a closed form for the correlations between the microvariables in terms of the temperature of the heat bath as a power law. From this, we obtain an information geometric method for studying Hamiltonian dynamics in the canonical ensemble. We illustrate the results with two examples: a pair of interacting harmonic oscillators presenting phase transitions and the 2×2 Gaussian ensembles. In both examples the scalar curvature results a global indicator of the dynamics.

© 2017 Published by Elsevier B.V.

1. Introduction

The possibility of that relevant features of the dynamics of a system could be obtained from the differential geometric structure of the probabilities distributions gave place to the first encounter between differential geometry and probability theory. In this sense, geometrization of thermodynamics and statistical mechanics constituted the most important achievement in the subject with several approaches like the obtained by means of the internal energy as considered Weinhold [1], or the Ruppeiner metric given by the second moments of thermodynamical fluctuations [2], among others. With the same Riemannian character of these approaches, another formulations based in the thermodynamic of parameters were established in the field of the statistical mechanics like the foundational works of Rao [3] and Amari [4], and also given by Ingarden [5], Janyszek [6]. The successful application of all this vast body of approaches in characterizing several phenomena, such as the phase transitions and their critical points in non ideal gases, gave them entity to constitute a

---

discipline within the information theory, called *Information Geometry*. Curved statistical manifolds are the subject of study of the information geometry, and they have associated the Fisher–Rao metric [3] which in turn is linked to the concepts of entropy and Fisher information. Generalized extensions of the information geometry [7–9] with regard to nonextensive formulation of statistical mechanics [10] has been also considered. The utility of information geometry is not only limited to thermodynamics and statistical mechanics. For instance, it has been applied in quantum mechanics leading a quantum generalization of the Fisher metric [11], and recently also in nuclear plasmas [12]. In particular, an application of information geometry to chaos can be performed by considering complexity on curved manifolds [13–16]. In this approach, asymptotic expressions for information measures are obtained by means of geodesic equations leading to a *criterion* for characterizing global chaos on statistical manifolds [16]: the more negative is the curvature, the more chaotic is the dynamics. As usual, chaos can be characterized in terms of diverging initially nearby trajectories [17]. For the statistical models this condition results in the divergence of geodesic paths on the statistical manifold and constitutes a local criterion for chaos.

Besides, in dynamical systems theory, the *ergodic hierarchy* (EH) characterizes the chaotic behavior in terms of a type of correlation between subsets of the phase space [18,19]. In the asymptotic limit of large times, the EH establishes that the dynamics is more chaotic when the correlation decays faster. According to the correlation decays, the four levels of EH are, from the weakest to the strongest: ergodic, mixing, Kolmogorov, and Bernoulli. In particular, in mixing systems any two subsets enough separated in time can be considered as "statistically independent", which allows one to use a statistical description of the behavior of the system. In addition, in quantum chaos, the statistical independence is present in the universal statistical properties of energy levels which are given by the Gaussian ensembles [20–23]. In Gaussian ensembles theory one assumes that in a fully chaotic quantum system the interactions are neglected in such way that the Hamiltonian matrix elements can be considered statistically independent [24]. Related to this, in [25–27] a quantum extension of the EH was proposed, called the *quantum ergodic hierarchy*, which allowed to provide a characterization of the chaotic behaviors of the Casati–Prosen model [23] and the kicked rotator [20–22].

Inspired by the characterizations of quantum chaotic systems made in [25–29] and making use of curved statistical models, we propose a generalization of the ergodic, mixing and Bernoulli levels of the EH in the context of the information geometry, which we called *Information Geometric Ergodic Hierarchy* (IGEH). In order to use it, we define a distinguishability measure for a 2*D* correlated model that allows us to give an upper bound for the correlation of IGEH. Moreover, considering Hamiltonian systems belonging to the canonical ensemble we also give a method for characterizing their dynamics in terms of the statistical parameters and the levels of the IGEH.

In this way, our main contribution is two-fold: (1) an intimately connection between statistical models and probability distributions of the canonical ensemble which allows one to geometrize the phase transitions, and (2) an information geometric version of the ergodic hierarchy as an alternative framework for studying the chaotic dynamics in curved statistical models.

The paper is organized as follows. In Section 2, we give the notions and concepts of information geometry used throughout the paper, along with brief description of a 2*D* correlated model. In Section 3, we make a brief review of the levels the ergodic hierarchy. Section 4 is devoted to an information geometric definition of the ergodic hierarchy by expressing the correlations in terms of probability distributions instead of subsets of phase space. Next, in Section 5 we define a distinguishability measure for the 2*D* correlated model and an upper bound for the correlation of the IGEH is given. In Section 6, we establish the connection between the statistical models and the family of probability distributions belonging to the canonical ensemble. For quadratic Hamiltonian systems we show that their associated statistical models are the multivariate Gaussian ones, for which we obtain a closed form for the covariance matrix determinant in terms of the Hessian matrix of the Hamiltonian. Here we also give an upper bound for the IG correlation where the temperature of the heat bath is considered as an external parameter. In Section 7, we illustrate the formalism with two examples: a pair of interacting harmonic oscillators presenting phase transitions in the canonical ensemble and the 2×2 Gaussian Orthogonal Ensemble (GOE). Also, a panoramic outlook of the IGEH is sketched. Finally, in Section 8 we draw some conclusions, and future research directions are outlined.

2. Elements of information geometry

We begin by introducing some fundamentals and concepts, following the definitions given in [4].

2.1. Statistical models

Given an abstract set X one can consider the set M of all the probability density functions (PDFs) p defined on X, i.e.

$$M = \left\{ p : \ p : X \to \mathbb{R}, \ p(x) \geq 0, \ \int_X p(x)dx \right\} \tag{1}$$

where $\mathbb{R}$ is the set of real numbers and the integration must be replaced by a sum when X is discrete. Consider a subset $S \subset M$ such that each element of S may be parameterized using a m-real vector $\theta = (\theta_1, \ldots, \theta_m)$ so that

$$S = \{ p_\theta \in M : p_\theta = p(x; \theta), \ \theta = (\theta_1, \ldots, \theta_m) \in \Theta \} \tag{2}$$

where Θ is a subset of $\mathbb{R}^m$. Then, if the mapping $\theta \mapsto p_\theta$ is injective it is said that S is a *statistical model* on X. The dimension of the statistical model is that of the macrospace, i.e. it is m-dimensional. The physical interpretation of X and Θ is as follows.

Generally, X represents the microscopic variables of the system under study which typically are difficult to control, for instance the positions of all the particles in a gas. Thus, X is called the *microspace* and x are the microvariables. On the other hand, Θ represent the macroscopic variables that can be measured in an experiment, like the mean value or the moments of the microvariables. It is said that Θ is the *macrospace* and $\theta_1, \ldots, \theta_m$ are the macrovariables. Since the microspace is fixed by the system then one can only choose the macrospace, and in this way the statistical model is established. Then, the statistical models are system-specific, from which follows that a statistical model could be useful for a system while that for another not.

2.2. Metric structure of the statistical manifold

Next step is to describe the behavior of a system by means of a previously and adequately chosen, statistical model. For this, some kind of dynamics must be introduced on the statistical model. In information geometry this is accomplished by means of the Fisher–Rao tensor

$$g_{ij}(\theta) = \int_X dx\, p(x;\theta) \frac{\partial \log p(x;\theta)}{\partial \theta_i} \frac{\partial \log p(x;\theta)}{\partial \theta_j} \quad i,j = 1, \ldots, m \tag{3}$$

where $p(x;\theta)$ is a generic element of S. The metric tensor g_{ij} endows the dynamics to the macrospace in terms of the geodesic equations for the macrovariables θ_i, i.e. S results to be a statistical manifold. More precisely, S is a Riemannian manifold and the statistical character is due to the elements of S are probability distributions.

Thus, the main goal of the statistical models is to obtain some relevant information about the dynamics by means of the geodesic equations and geometrical quantities like the Ricci tensor, the scalar curvature etc. In this sense, we will use a criteria given by Cafaro et al. [13–15] to characterize global chaos on statistical models: "*the more negative is the scalar curvature, the more chaotic is the dynamics*". From the metric tensor (3) one can obtain the geodesic equations for the macrovariables $\theta_1, \ldots, \theta_m$ along with the following geometrical quantities that we will use throughout the paper.

$$\text{Geodesic equations}: \quad \frac{d\theta_k}{d\tau} + \Gamma_{ij}^k \frac{d\theta_i}{d\tau} \frac{d\theta_j}{d\tau} = 0, \quad \forall\, k = 1, \ldots, m \tag{4}$$

$$\text{Christoffel symbols}: \quad \Gamma_{ij}^k = \frac{1}{2} g^{im} \left(g_{mk,l} + g_{ml,k} - g_{kl,m} \right) \tag{5}$$

$$\text{Riemann curvature tensor}: \quad R_{iklm} = \frac{1}{2} \left(g_{im,kl} + g_{kl,im} - g_{il,km} - g_{km,il} \right) + g_{np} \left(\Gamma_{kl}^n \Gamma_{im}^p - \Gamma_{km}^n \Gamma_{il}^p \right) \tag{6}$$

$$\text{Ricci tensor}: \quad R_{ik} = g^{lm} R_{limk} \tag{7}$$

$$\text{Scalar curvature}: \quad R = g^{ik} R_{ik} \tag{8}$$

where the comma in the subindexes denotes the partial derivative operation (of first and second orders), g^{kl} is the inverse of g_{ij}, and τ is a scalar that parameterizes the geodesic curves.

2.3. The 2D correlated Gaussian model

Most statistical models used in the literature are the so called *Gaussian models*, due to its wide versatility for describing multiple phenomena. This models are obtained by choosing the subfamily S as the set of multivariate Gaussian distributions. If $(x_1, \ldots, x_n) \in \mathbb{R}^n$ are the microvariables and there is no correlations between them, then $(\mu_1, \ldots, \mu_n, \sigma_1, \ldots, \sigma_n) \in \mathbb{R}^n \times \mathbb{R}_+^n$ are the set of macrovariables, where μ_i and σ_i^2 correspond to the mean value and the variance of the ith microvariable. However, for a more realistic description of the system the correlations between each of the microvariables must be taken into account. Considering the family of bivariate (binormal) distributions, one of the Gaussian models of lower dimension that present correlations can be obtained, which is given by

$$p(x, y; \mu_x, \mu_y, \sigma_x, \sigma_y, r) = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1 - r^2}}$$

$$\times \exp\left(-\frac{1}{2(1 - r^2)} \left[\frac{(x - \mu_x)^2}{\sigma_x^2} + \frac{(y - \mu_y)^2}{\sigma_y^2} - \frac{2r(x - \mu_x)(y - \mu_y)}{\sigma_x \sigma_y} \right] \right) \tag{9}$$

where $\sigma_{xy} = r\sigma_x\sigma_y$ is the covariance between x and y and r is the correlation coefficient that assumes values within the ranges $-1 \leq r \leq 1$. Here the microspace is $X = \{(x, y) \in \mathbb{R}^2\}$ and the macrospace is $\Theta = \{(\mu_x, \mu_y, \sigma_x, \sigma_y) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+\}$. Nevertheless, one still can obtain a non trivial description by adding the following *macroscopic constraint*

$$\Sigma^2 = \sigma_x \sigma_y \tag{10}$$

where Σ is a constant belonging to $\mathbb{R}_+$. Mathematically, the effect of Σ is to restrict the dynamics to the submanifold $\Theta \cap \{\Sigma = \sigma_x \sigma_y\}$. Physically, Σ resembles the minimum uncertainty relation when one chooses x as the position of a particle

and y its conjugate variable. Moreover, this interpretation allows to give an explanation of the phenomenon of "suppression of classical chaos by quantization" from an information geometric point of view [16]. For the sake of simplicity one also can fix the mean value μ_y of y as zero. With the help of (10) then one can rewrite (9), thus obtaining the non-trivial correlated statistical model of lower dimensionality

$$p(x, y; \mu_x, \sigma, r) = \frac{1}{2\pi \Sigma^2 \sqrt{1-r^2}} \exp\left(-\frac{1}{2(1-r^2)} \left[\frac{(x-\mu_x)^2}{\sigma^2} + \frac{y^2\sigma^2}{\Sigma^4} - \frac{2r(x-\mu_x)y}{\Sigma^2}\right]\right) \tag{11}$$

where we renamed σ_x as σ. This is the so called *the 2D correlated model* [16], since it present correlations between x and y by means of $\Sigma^2 = \sigma_x\sigma_y$ and its macrospace $\Theta = \{(\mu_x, \sigma) \in \mathbb{R} \times \mathbb{R}_+\}$ is bidimensional. It should be noted that here r is considered as an external parameter that does not belong to the macrospace.

Two remarks about the Eq. (11) deserve to be made. First, the assumption $\mu_y = 0$ would seem to be an asymmetry in the choice of the macrovariables of the macrospace. However, this is not the case since considering $\mu_x = 0$ one gets the same model, i.e. endowed with the same physical characteristics, due to the symmetry of the crossed term between x and y within the exponential of (11). Second, the hypothesis of considering $\mu_x = 0$ or $\mu_y = 0$ is with the aim of obtaining the minimal and relevant model presenting correlations between x and y.

For this model, from Eqs. (3)–(8) one can obtain the Fisher tensor along with the following geometrical quantities.

$$\text{Fisher–Rao metric}: \ g_{ij}(\theta) = \begin{pmatrix} \dfrac{1}{\sigma^2(1-r^2)} & 0 \\ 0 & \dfrac{4}{\sigma^2(1-r^2)} \end{pmatrix} \tag{12}$$

$$\text{Non-vanishing Christoffel symbols}: \ \Gamma_{12}^1 = \Gamma_{21}^1 = -\frac{1}{\sigma}, \qquad \Gamma_{11}^2 = \frac{1}{4\sigma}, \qquad \Gamma_{22}^2 = -\frac{1}{\sigma} \tag{13}$$

$$\text{Non-vanishing Ricci tensor components}: \ R_{11} = -\frac{1}{4\sigma^2}, \qquad R_{22} = -\frac{1}{\sigma^2} \tag{14}$$

$$\text{Scalar curvature}: \ R(r) = -\frac{1}{2}(1-r^2), \quad -1 \leq r \leq 1. \tag{15}$$

From (15) one can see that the curvature has a minimum value $R = -\frac{1}{2}$ when $r = 0$ (absence of correlations) while for $|r| \to 1$ (maximally correlated case) it has a maximum value $R = 0$. In terms of the criterium of global chaos this can be interpreted as: the dynamics of the uncorrelated case is more chaotic than the corresponding to the maximally correlated case. Moreover, the divergence of the metric observed for $|r| \to 1$ expresses the maximally correlated case as a critical point of the dynamics.

3. The ergodic hierarchy

In classical chaos, the exponential instability implies continuous spectrum, and therefore, a decay of correlations in such a way that for large times the measure of the intersection between two sets of phase space (separated from each other in time) tends to the product of their measures. This is the well known *mixing* property, and constitutes one of the foundations of the statistical mechanics. The main feature of mixing is that it establishes the statistical independence of different parts of a trajectory, when sufficiently separated in time. This is the main reason for the application of probability theory in the classical domain, which allows one to calculate statistical features such as diffusion, relaxation and distribution functions [22]. Consequently, the description in terms of trajectories can be replaced by an equivalent one in terms of distribution functions, which, if not singular, represent not a single trajectory but a continuum of them.

In ergodic theory, any classical system is represented mathematically by a dynamical system $(X, \Sigma, \mu, \{T_t\}_{t\in J})$ where X is a set, Σ is a sigma-algebra of X, μ a measure defined over Σ and $\{T_t\}_{t\in J}$ a group of measure-preserving transformations. The ergodic hierarchy ranks the chaos of a dynamical system according to a type of correlation $C(T_tA, B)$ between two subsets A and B of X that are separated by a time t. This is defined as [18,19]

$$C(T_tA, B) = \mu(T_tA \cap B) - \mu(A)\,\mu(B). \tag{16}$$

The ergodic, mixing and Bernoulli levels of the EH are given in terms of (16) in the following way. Given two arbitrary sets $A, B \in X$, it is said that T_t is

- *ergodic* if

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T C(T_tA, B)\, dt = 0, \tag{17}$$

- *mixing* if

$$\lim_{t\to\infty} C(T_tA, B) = 0, \tag{18}$$

- *Kolmogorov* if for all integer r, for all $A_0, A_1, \ldots, A_r \subseteq X$, and for all $\varepsilon > 0$ there exists a positive integer $n_0 > 0$ such that if $n \geq n_0$ one has

$$|C(A_0, B)| < \varepsilon, \quad \text{for all } B \in \sigma_{n,r}(A_1, \ldots, A_r) \tag{19}$$

where $\sigma_{n,r}(A_1, \ldots, A_r)$ is the minimal σ-algebra generated by $\{T^k A_i : k \geq n \; ; \; i = 1, \ldots, r\}$.
- *Bernoulli* if

$$C(T_t A, B) = 0 \quad \text{for all } t \geq 0. \tag{20}$$

In ergodic systems the correlation vanishes "in time average" for large times while in mixing systems $C(T_t A, B)$ vanishes for $t \to \infty$. In Kolmogorov systems the correlations between an arbitrary set and another one belonging to the σ-algebra $\sigma_{n,r}(A_1, \ldots, A_r)$ cancel for $n \to \infty$. In Bernoulli systems the correlation is zero for all times. These levels classify the dynamics according to Eqs. (17), (18), (19), and (20), from the weakest level (the ergodic) to the strongest (the Bernoulli). The following strict inclusions hold:

$$\text{ergodic} \supset \text{mixing} \supset \text{Kolmogorov} \supset \text{Bernoulli}.$$

In order to express $C(T_t A, B)$ by means of probability distributions it is more convenient to use the definition (16) in terms of distribution functions, which is given by [19]

$$C(f \circ T_t, g) = \int_X (f \circ T_t)(x) g(x) dx - \int_X f(x) dx \int_X g(x) dx \quad \forall f, g \in \mathbb{L}^1(X) \tag{21}$$

where $f \circ T_t$ denotes the composition of f and T_t, i.e. $f \circ T_t(x) = f(T_t(x))$ for all $x \in X$ and now the role of A, B is played by the functions $f, g \in \mathbb{L}^1(X)$. Physically, f represents any initial density function of the classical system whose value at time t is given by $f \circ T_t$ with T_t the classical Liouville evolution (in Hamiltonian systems).

4. An information geometric version of the ergodic hierarchy

Following the idea of characterizing chaos by means of the ergodic hierarchy [25–27], now we consider an extension of the EH within the context of the information geometry. In information geometry one has probability distributions p_θ that depend on a set of parameters θ, and the dynamics of the macrovariables θ is performed along the geodesics of the statistical manifold. Moreover, in the statistical manifold the role of time variable t of dynamical systems is played by a parameter τ along the geodesics.

In order to introduce the tools of information geometry, we propose the following approach by defining a correlation between functions as the macrovariables θ evolve along the geodesics. Given N functions $f(x_i)$, each one of them in terms of the variable x_i for all $i = 1, \ldots, N$, we define the *information geometric correlation* (IG correlation) $C(f_1, \ldots, f_N, \tau)$ between $f_1, \ldots, f_N$ at time-like parameter τ as

$$C(f_1, \ldots, f_N, \tau) \doteq \int p(x_1, \ldots, x_N; \theta(\tau)) f_1(x_1) \cdots f_N(x_N) dx_1 \cdots dx_N - \prod_{i=1}^{N} \int p_i(x_i; \theta(\tau)) f_i(x_i) dx_i \tag{22}$$

where $\theta(\tau) = (\theta_1(\tau), \ldots, \theta_M(\tau))$ is the M-dimensional vector of the macrovariables at "time" τ and,

$$p_i(x_i; \theta(\tau)) = \int p(x_1, \ldots, x_N; \theta(\tau)) \prod_{j \neq i} dx_j, \quad i = 1, \ldots, N \tag{23}$$

are the marginal distributions of $p(x_1, \ldots, x_N; \theta(\tau))$. From (22) one can see that $C(f_1, \ldots, f_N, \tau)$ measures how independent the variables $x_1, \ldots, x_N$ are at time-like τ. This can be considered as a sort of information geometric generalization of the EH correlation.

Having established $C(f_1, \ldots, f_N, \tau)$ and taking into account the ergodic, mixing and Bernoulli levels given by Eqs. (17), (18) and (20), we define the *information geometric ergodic hierarchy* (IGEH) as follows. For the sake of simplicity and since we focus mainly on the ergodicity and mixing properties (which are the fundamentals of statistical mechanics), in this contribution we do not extend the Kolmogorov level that involves a σ-algebra.[1] Given a set of N arbitrary functions $f_1(x_1), \ldots, f_N(x_N)$ we say that now the statistical model is

- *IG ergodic* if

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T C(f_1, \ldots, f_N, \tau) d\tau = 0, \tag{24}$$

---

[1] In fact, the extension of the Kolmogorov level requires a more careful treatment than the rest of the levels, due to the presence of the σ-algebra $\sigma_{n,r}(A_1, \ldots, A_r)$, that is out of the technical scope of this work.

- *IG mixing* if

$$\lim_{\tau \to \infty} C(f_1, \ldots, f_N, \tau) = 0,$$ (25)

- *IG Bernoulli* if

$$C(f_1, \ldots, f_N, \tau) = 0 \quad \text{for all } t \in \mathbb{R}.$$ (26)

As in the ergodic hierarchy, the following strict inclusions hold:

IG ergodic $\supset$ IG mixing $\supset$ IG Bernoulli.

For instance, a statistical model that is IG ergodic can be given by assuming that $C(f_1, \ldots, f_N, \tau)$ is proportional to $\sin(\alpha \tau) \|f_1\|_1 \ldots \|f_N\|_1$ with $\alpha \in \mathbb{R}$. Making this replacement in (24) one obtains that $\lim_{T \to \infty} \frac{1}{T} \int_0^T C(f_1, \ldots, f_N, \tau) d\tau$ is equal to zero. Since $\sin(\alpha \tau) \|f_1\|_1 \ldots \|f_N\|_1$ oscillates, then this model cannot be IG mixing nor IG Bernoulli. Examples of statistical models that are IG mixing and IG Bernoulli will be illustrated in Section 7. Our approach, thus, deals with ergodic hierarchy in statistical models from an information geometry point of view.

5. A measure of distinguishability for the 2*D* correlated model

In order to use the levels of the IGEH for characterizing the dynamics of statistical models one should have a manner of determining the decay of the correlation $C(f_1, \ldots, f_N, \tau)$ in Eqs. (24), (25) or (26). For the family of the 2*D* correlated probabilities $p(x, y | \mu_x, \sigma; r)$ of (11), we define a distinguishability measure $F : \{p(x, y; \mu, \sigma, r) \mid \mu_x \in (-\infty, \infty), \ \sigma \in (0, \infty), \ -1 \leq r \leq 1\} \longmapsto \mathbb{R}$, given by

$$F(p) \doteq \|p(x, y; \mu_x, \sigma, r) - p_1(x)p_2(y)\|_\infty = \max_{(x,y) \in \mathbb{R}^2} |p(x, y; \mu_x, \sigma, r) - p_1(x)p_2(y)|$$ (27)

where $p_1(x), p_2(y)$ are the marginal distributions of $p(x, y; \mu_x, \sigma, r)$. Furthermore, if $f_1(x), f_2(y) \in \mathbb{L}^1(\mathbb{R})$ are arbitrary functions of x and y, then we have

$$|C(f_1, f_2, \tau)| = \left| \int_{\mathbb{R}^2} p(x, y; \mu_x, \sigma, r) f_1(x) f_2(y) dx dy - \int_{\mathbb{R}} p_1(x) f_1(x) dx \int_{\mathbb{R}} p_2(y) f_2(y) dy \right|$$

$$\leq \left\{ \max_{(x,y) \in \mathbb{R}^2} |p(x, y; \mu_x, \sigma, r) - p_1(x)p_2(y)| \right\} \left| \int_{\mathbb{R}^2} dx dy f_1(x) f_2(y) \right| \leq F(p) \|f_1 f_2\|_1.$$ (28)

Eq. (28) expresses that $F(p) \|f_1 f_2\|_1$ is an upper bound for $|C(f_1, f_2, \tau)|$. Therefore, it is convenient to find an analytic expression for (27). After some algebra one can obtain that[2]

$$F(p) = |r| \left(\sqrt{1 - r^2}(1 + |r|) \right)^{-1 - \frac{1}{|r|}} \quad \text{for all } r \in [-1, 1].$$ (29)

The behavior of $F(p)$, which is independent of μ_x and σ, is shown in Fig. 1. Two relevant regions, corresponding to the limiting cases $r \to 0$ and $r \to \pm 1$, can be well distinguished. The region $r \to 0$ corresponds to the zone where the statistical model is characterized by the IG mixing and IG Bernoulli levels, with the particularity that the variables of microspace are uncorrelated. Moreover, one can see that near to $r = 0$ the decay is linear in r. The curve $F(p)$ also shows that, if $r \to 0$ when $\tau \to \infty$, then the statistical model is IG mixing.

In the region $r \to \pm 1$ the measure $F(p)$ diverges corresponding to the maximally correlated case, which physically means that the system presents strong correlations between the variables of microspace. Due to the correlations are strong in this regime the statistical model cannot be IG mixing nor IG Bernoulli.

Finally, it should be noted that $F(p)$ does not allow one to distinguish between two probability distributions having r and $-r$ respectively. The symmetry respect to the axis $r = 0$ is due to the mathematical form of the infinite norm $\|.\|_\infty$ in the definition (27). That is, with other choices of $F(p)$ one could distinguish states (probability distributions) with correlation coefficients r and $-r$.

6. Geometrizing the canonical ensemble by means of statistical models and the IGEH

On the basis of the above characterization of the dynamics of the macrospace in terms the IGEH, our next aim is to give a method for studying the dynamics of a system belonging to the canonical ensemble.

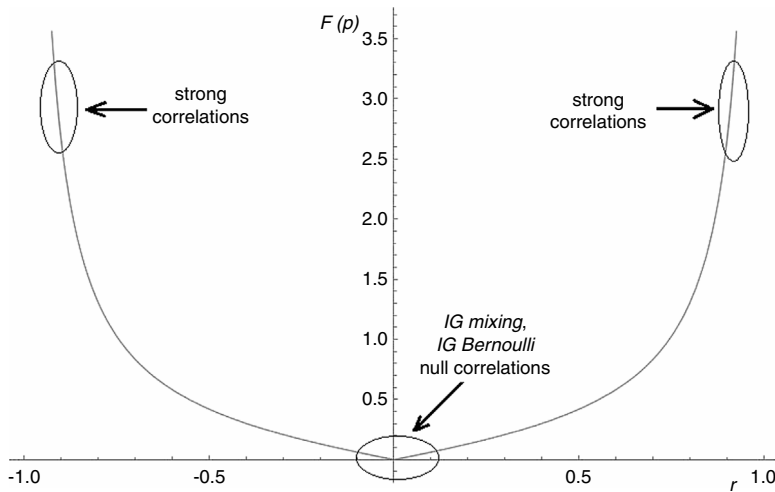[2] The demonstration can be found in the Appendix.

Fig. 1. Behavior of $F(p)$ in terms of the correlation coefficient. Near to the region $r = 0$ the statistical model belongs to the IG mixing level while for $r = 0$ it is IG Bernoulli. When $r \rightarrow \pm 1$ one has that $F(p)$ diverges with the presence of strong correlations. The discontinuity of the slope of F at $r = 0$ is due to the particular form of $F(p)$, i.e. the maximum operation of infinite norm can present discontinuities in its derivatives.

6.1. Canonical ensembles in the context of the information geometry

Beyond the relationship between statistical models and statistical physics has been already established [4,30], extensions in several directions have been recently introduced [31–38], with a particular focusing on the exponential families since they represent mathematically the Liouville densities of the statistical ensembles. Relevant consequences from these researches such as the connection between Hessian structures and exponential families [37], nonextensive statistical models [33–35,38] and other extensions [31,32,36] has been reported.

In order to apply the IGEH to the canonical ensemble, here we obtain some explicit formulas that relate the statistical parameters of multivariate Gaussian distributions (which are a special case of exponential family) with the physical parameters of quadratic Hamiltonians. In the present contribution we only focus on the multivariate Gaussian distributions and we will omit the definition of exponential families of a more general character. We begin by considering the family of probability density functions given by the classical canonical ensemble

$$p(q, p; \theta) = \frac{1}{Z(\theta)} \exp \{-\beta E(q, p, \theta)\} \tag{30}$$

where $Z(\theta) = \int p(q, p; \theta) dq dp$ is the well known partition function, $\beta = \frac{1}{k_B T}$ is the Boltzmann factor, $E(q, p, \theta)$ is the energy of the system expressed in terms of the phase space coordinates $(q, p) \in \Gamma$ (with Γ the phase space), and θ are the macrovariables of the system. The function $p(q, p; \theta)$ represents the probability density of the microstate (q, p) of the system, corresponding to the macrostate given by the macroscopic parameters θ, when contacted in thermal equilibrium with a heat bath at a fixed temperature T. In this way, one can see that there is a biunivocal correspondence between statistical models and statistical ensembles: each pair of microvariables (q, p) corresponds to a microstate of the ensemble, and each value of the macrovariable θ is associated to a macrostate of the system.

Assuming a $2n$-dimensional phase space Γ and a m-dimensional macrospace Θ, from (30) and using (3) one obtains the Fisher tensor for the canonical ensemble

$$
\begin{aligned}
g_{ij}(\theta)_{CE} &= \frac{\beta^2}{Z(\theta)} \int dq dp \exp \{-\beta E(q, p, \theta)\} \frac{\partial E(q, p; \theta)}{\partial \theta_i} \frac{\partial E(q, p; \theta)}{\partial \theta_j} \\
&= \beta^2 \left\langle \frac{\partial E(q, p; \theta)}{\partial \theta_i} \frac{\partial E(q, p; \theta)}{\partial \theta_j} \right\rangle_{CE} \quad i, j = 1, \ldots, m
\end{aligned}
\tag{31}
$$

where q and p are a short notation for $(q_1, \ldots, q_n)$ and $(p_1, \ldots, p_n)$, and the suffix CE stands for the canonical ensemble. The formula (31) is the expression of the canonical ensemble in the context of the information geometry. All the dynamics over the macrospace can be derived by means of the Eqs. (4)–(8), thus obtaining a geometrical characterization of the canonical ensemble.

6.2. Dynamics of the canonical ensemble in terms of the IGEH levels

The particular dependence of the energy $E(q, p; \theta)$ on the microvariables (q, p) determines the correlations between them, which are reflected in the probability distribution $p(q, p; \theta)$. Since the IG correlation $C(f_1, \ldots, f_N, \tau)$ measures the

degree of statistical independence between the microvariables of the microspace at time-like parameter τ, then for a given expression of the energy it is desirable to know what is the form of $C(f_1, \ldots, f_N, \tau)$. We focus on a particular form of the energy, i.e. when $E(q, p; \theta)$ is a quadratic function of (q, p). The physical relevance of this assumption lies, among other things, in the fact that it allows one to study the dynamics of the system near of an equilibrium point. In this case, the formula of the energy is given by

$$E(q, p; \theta) = (q - q_0, p - p_0) \, \mathbf{G} \, (q - q_0, p - p_0)^T$$

$$\mathbf{G}_{ij} = \frac{1}{2} \frac{\partial^2 E}{\partial x_i \partial x_j}(q_0, p_0), \quad i, j = 1, \ldots, n \tag{32}$$

$$(x_1, \ldots, x_n, x_{n+1}, \ldots, x_{2n}) = (q_1, \ldots, q_n, p_1, \ldots, p_n)$$

with $\mathbf{G}$ the Hessian of E around at a point (q_0, p_0) and $(q - q_0, p - p_0)^T$ the transposed of $(q - q_0, p - p_0)$. Then, the probability distribution $p(q, p; \theta)$ adopts the form

$$p(q, p; \theta) = \frac{1}{Z(\theta)} \exp\left\{-\beta(q - q_0, p - p_0) \, \mathbf{G} \, (q - q_0, p - p_0)^T\right\} \tag{33}$$

with $Z(\theta)^{-1}$ the normalization factor. From (33) one can see that $p(q, p; \theta)$ is nothing but a $2n$-multivariate Gaussian distribution on the microvariables $q_1, \ldots, q_n, p_1, \ldots, p_n$. At this point it is convenient to recall the expression of the $2n$-multivariate Gaussian distribution $p(x; \mu, \mathbf{\Sigma})$

$$p(x; \mu, \mathbf{\Sigma}) = \left(\frac{1}{2\pi}\right)^n \frac{1}{\sqrt{|\mathbf{\Sigma}|}} \exp\left\{-\frac{1}{2}(x - \mu) \, \mathbf{\Sigma}^{-1} \, (x - \mu)^T\right\} \tag{34}$$

with $x = (x_1, \ldots, x_{2n})$ the vector of microvariables, $\mu = (\mu_1, \ldots, \mu_{2n})$ the mean value vector, $\mathbf{\Sigma}$ the covariance matrix and $\mathbf{\Sigma}$ its inverse, and $|\mathbf{\Sigma}|$ the determinant of $\mathbf{\Sigma}$. Moreover, the marginals $p_i(x_i; \mu_i, \mathbf{\Sigma}_{ii})$ of $p(x; \mu, \mathbf{\Sigma})$ are given by

$$p_i(x_i; \mu_i, \mathbf{\Sigma}_{ii}) = \frac{1}{\sqrt{2\pi \mathbf{\Sigma}_{ii}}} \exp\left\{-\frac{1}{2} \frac{1}{\mathbf{\Sigma}_{ii}}(x_i - \mu_i)^2\right\}, \quad i = 1, \ldots, 2n \tag{35}$$

where $\mathbf{\Sigma}_{ij}$ stands for the ijth matrix element of $\mathbf{\Sigma}$ with $i, j = 1, \ldots, 2n$.

Let us consider $2n$ arbitrary functions $f_1(x_1), \ldots, f_{2n}(x_n) \in \mathbb{L}^1(\mathbb{R})$. Then, replacing (34) and (35) in (22) one has that

$$C(f_1, \ldots, f_{2n}, \tau) \doteq \int_{\mathbb{R}^{2n}} \left[\left(\frac{1}{2\pi}\right)^n \frac{1}{\sqrt{|\mathbf{\Sigma}|}} \exp\left\{-\frac{1}{2}(x - \mu) \, \mathbf{\Sigma}^{-1} \, (x - \mu)^T\right\} f_1(x_1) \cdots f_{2n}(x_{2n})\right] dx_1 \cdots dx_{2n}$$

$$- \prod_{i=1}^{2n} \int_{\mathbb{R}} \left[\frac{1}{\sqrt{2\pi \mathbf{\Sigma}_{ii}}} \exp\left\{-\frac{1}{2} \frac{1}{\mathbf{\Sigma}_{ii}}(x_i - \mu_i)^2\right\} f_i(x_i)\right] dx_i \tag{36}$$

is the IG correlation for the family of $2n$-multivariate Gaussian distributions. As in the $2D$ correlated model, one can obtain an upper bound for $C(f_1, \ldots, f_{2n}, \tau)$ as follows.

$$|C(f_1, \ldots, f_{2n}, \tau)|$$

$$\leq \max_{x \in \mathbb{R}^{2n}} \left\{\left(\frac{1}{2\pi}\right)^n \frac{1}{\sqrt{|\mathbf{\Sigma}|}} \exp\left\{-\frac{1}{2}(x - \mu) \, \mathbf{\Sigma}^{-1} \, (x - \mu)^T\right\} - \prod_{i=1}^{2n} \frac{1}{\sqrt{2\pi \mathbf{\Sigma}_{ii}}} \exp\left\{-\frac{1}{2} \frac{1}{\mathbf{\Sigma}_{ii}}(x_i - \mu_i)^2\right\}\right\}$$

$$\times \left|\int_{\mathbb{R}^{2n}} f_1(x_1) \cdots f_{2n}(x_{2n}) dx_1 \cdots dx_{2n}\right|$$

$$\leq \max_{x \in \mathbb{R}^{2n}} \left\{\left(\frac{1}{2\pi}\right)^n \frac{1}{\sqrt{|\mathbf{\Sigma}|}} \exp\left\{-\frac{1}{2}(x - \mu) \, \mathbf{\Sigma}^{-1} \, (x - \mu)^T\right\} - \prod_{i=1}^{2n} \frac{1}{\sqrt{2\pi \mathbf{\Sigma}_{ii}}} \exp\left\{-\frac{1}{2} \frac{1}{\mathbf{\Sigma}_{ii}}(x_i - \mu_i)^2\right\}\right\}$$

$$\times \|f_1(x_1) \cdots f_{2n}(x_{2n})\|_1. \tag{37}$$

This inequality expresses an upper bound of the IG correlation for the family of the $2n$-multivariate Gaussian distributions, where the maximum is a measure of distinguishability, and the parameters $\mu = \mu(\tau)$, $\mathbf{\Sigma} = \mathbf{\Sigma}(\tau)$ are dependent on τ along the geodesics by means of the application of Eqs. (3)–(5) to $p(x; \mu, \mathbf{\Sigma})$.

Now we can set the upper bound of (37) in the language of the canonical ensemble. By simple inspection of Eqs. (32)–(35), if one makes the following replacements

$$x = (x_1, \ldots, x_{2n}) \longrightarrow (q, p) = (q_1, \ldots, q_n, p_1, \ldots, p_n)$$

$$\mu = (\mu_1, \ldots, \mu_{2n}) \longrightarrow (q_0, p_0) = (q_{10}, \ldots, q_{n0}, p_{10}, \ldots, p_{n0})$$

$$|\mathbf{\Sigma}|^{-1} \longrightarrow |2\beta \mathbf{G}_{ij}| = (2\beta)^{2n}|\mathbf{G}_{ij}| \tag{38}$$

$$\mathbf{\Sigma}_{ij}^{-1} \longrightarrow 2\beta \mathbf{G}_{ij} = \beta \frac{\partial^2 E}{\partial x_i \partial x_j}(q_0, p_0), \quad i, j = 1, \ldots, 2n$$

in (37) then one obtains

$$
\begin{aligned}
&|C(f_1, \ldots, f_{2n}, \tau)_{CE}| \\
&\quad \le \max_{(q,p) \in \Gamma} \left[\left(\frac{\beta}{\pi} \right)^n \frac{1}{\sqrt{|\mathbf{G}^{-1}|}} \exp \left\{ -\beta (q - q_0, p - p_0)\, \mathbf{G}\, (q - q_0, p - p_0)^T \right\} \right. \\
&\qquad \left. - \prod_{i=1}^{n} \frac{\beta}{\pi \sqrt{(\mathbf{G}^{-1})_{ii}(\mathbf{G}^{-1})_{n+i n+i}}} \exp \left\{ -\beta \left(\frac{1}{(\mathbf{G}^{-1})_{ii}}(q_i - q_{i0})^2 + \frac{1}{(\mathbf{G}^{-1})_{n+i n+i}}(p_i - p_{i0})^2 \right) \right\} \right] \\
&\qquad \times \| f_1(q_1) \cdots f_n(q_n) f_{n+1}(p_1) \cdots f_{2n}(p_n) \|_1.
\end{aligned}
\tag{39}
$$

The inequality (39) is an upper bound of the IG correlation for a system in the canonical ensemble whose energy is a quadratic form, expanded around a point (q_0, p_0) of the phase space. Since the macrovariables (q_0, p_0) and $\mathbf{G}$ are functions of the time-like parameter τ then the utility of (39) is that, according to the way the IG correlation cancel for large values of τ, one can classify the statistical model as belonging to some of the levels of the IGEH. In this way, one can study phase transitions in terms of the IGEH levels as the macrovariables vary. For instance, if the dynamics in macrospace is such that the maximum in (39) tends to zero when $\tau \to \infty$, then one has that the canonical ensemble behaves as an IG mixing statistical model.

It should be noted that the multivariate Gaussian distribution only exists if the covariance matrix $\mathbf{\Sigma}$ is positive, which implies that $\mathbf{\Sigma}^{-1}$ must be also positive. Then, it follows that $\frac{\partial^2 E}{\partial^2 x_i}(q_0, p_0) > 0$. In particular, the stable equilibrium points of the system satisfy this condition.

Now we are in a position to reach one of our main contributions of this work. By the Eq. (38) and using the definition of the tensor $\mathbf{G}$, one finally obtains

$$
|\mathbf{\Sigma}| \left| \frac{\partial^2 E}{\partial x_i \partial x_j}(q_0, p_0) \right| = (k_B T)^{2n}.
\tag{40}
$$

This equation expresses an intimate relationship between the canonical ensemble and the statistical models, which one can express in words as follows:

> "*given a quadratic form of the energy, the determinant of the covariance matrix, which measures the correlations between the microvariables, is proportional to a power (the dimension of the phase space) of the temperature of the heat bath*".

Moreover, since the determinant of covariance matrix is a decreasing function of the correlations, as the temperature of the heat bath increases the correlations tend to be suppressed, as expected statistically. This result will be useful for characterizing phase transitions, as we shall see below.

7. Models and results

In order to illustrate the relevance of the IGEH we consider two examples belonging to different topics: an interacting bipartite system presenting phase transitions in the canonical ensemble, and the 2×2 Gaussian orthogonal ensemble. Next, we give a panoramic outlook of the IGEH and of the relationship between statistical models and Liouville densities of the canonical ensemble.

7.1. Phase transitions in a pair of interacting harmonic oscillators

Let us consider a pair of unidimensional and interacting harmonic oscillators in the canonical ensemble whose total energy is given by

$$
E(q_1, q_2, p_1, p_2; q_{10}, q_{20}, m_1, m_2, \omega_1, \omega_2, r) = T(p_1, p_2; m_1, m_2) + V(q_1, q_2; q_{10}, q_{20}, m_1, m_2, \omega_1, \omega_2, r)
$$

$$
T(p_1, p_2; m_1, m_2) = \frac{p_1^2}{2m_1} + \frac{p_2^2}{2m_2}
$$

$$
\begin{aligned}
V(q_1, q_2; q_{10}, q_{20}, m_1, m_2, \omega_1, \omega_2, r) &= \frac{1}{2} m_1 \omega_1^2 (q_1 - q_{10})^2 + \frac{1}{2} m_2 \omega_2^2 (q_2 - q_{20})^2 \\
&\quad - r \sqrt{m_1 m_2}\, \omega_1 \omega_2 (q_1 - q_{10})(q_2 - q_{20})
\end{aligned}
$$

where q_i, p_i, q_{i0}, m_i, and ω_i are the position, the momentum, the equilibrium position, the mass, and the frequency of the ith particle with $i = 1, 2$. Here $T(p_1, p_2; m_1, m_2)$ is the kinetic energy and $V(q_{10}, q_{20}, m_1, m_2, \omega_1, \omega_2, r)$ is the potential energy which is composed by three terms: the first two are the potential energy of each oscillator separately while the term $-r\sqrt{m_1 m_2}\, \omega_1 \omega_2 (q_1 - q_{10})(q_2 - q_{20})$ represents the interaction between the oscillators, and the coefficient $r \in [-1, 1]$ measures the strength coupling. For instance, when $r = 0$ the oscillators are uncoupled, and therefore, their motions are independent of each other.

Now, in order to use the analysis made about the 2*D* correlated model one must reduce the number of microvariables and macrovariables, and also impose some type of constraints. In this sense, we fix the masses m_1, m_2 and set $q_{20} = 0$. Also, we consider the following constraint

$$\Sigma^2 = \frac{k_B T_0}{\sqrt{m_1 m_2 \omega_1 \omega_2}} \tag{41}$$

where Σ is a real constant, k_B is the Boltzmann constant, and T_0 is a temperature of reference.[3] Here T_0 plays the role of being a temperature that breaks up the correlations between the microvariables. Two remarks about the constraint (41) deserve to be made. First, one can see that (41) corresponds to the constraint (10) in order to use the 2*D* correlated model. Second, the physical content of (41) is to introduce correlations between the positions q_1 and q_2 of the oscillators by restricting the dynamics to the submanifold $\Theta \cap \{\Sigma^2 = \sigma_1 \sigma_2\}$, with $\sigma_i = \sqrt{\frac{k_B T_0}{m_i \omega_i}}$ and $i = 1, 2$.

Due to the correlations are only between the position coordinates q_1, q_2 then one can reasonably neglect the momentum coordinates p_1, p_2 by integrating the probability distribution $p(q_1, q_2, p_1, p_2; \theta)$ given by the canonical ensemble, i.e. $\frac{1}{Z(\theta)} \exp\{-\beta E(q_1, q_2, p_1, p_2; \theta)\}$, over p_1 and p_2. And since only the kinetic energy has the dependence on p_1 and p_2 then this equivalent to consider a sort of marginal probability distribution with respect to the potential energy, i.e.

$$p(q_1, q_2; \theta) = A(\theta) \exp\{-\beta V(q_1, q_2; \theta)\}, \quad \text{with} \quad A(\theta)^{-1} = \int \exp\{-\beta V(q_1, q_2; \theta)\} dq_1 dq_2. \tag{42}$$

With the help of (41) one can express the potential energy as

$$V(q_1, q_2; q_{10}, \omega_1, r) = \frac{k_B T_0}{2} \left\{ \frac{(q_1 - q_{10})^2}{k_B T_0 (m_1 \omega_1^2)^{-1}} + \frac{q_2^2 k_B T_0 (m_1 \omega_1^2)^{-1}}{\Sigma^4} - \frac{2r(q_1 - q_{10})q_2}{\Sigma^2} \right\}. \tag{43}$$

Then, from (42) and (43) one has

$$p\left(q_1, q_2; q_{10}, \sqrt{k_B T_0 (m_1 \omega_1^2)^{-1}}, r\right) = A\left(q_{10}, \sqrt{k_B T_0 (m_1 \omega_1^2)^{-1}}, r\right)$$

$$\times \exp\left\{ -\beta \frac{k_B T_0}{2} \left\{ \frac{(q_1 - q_{10})^2}{k_B T_0 (m_1 \omega_1^2)^{-1}} + \frac{q_2^2 k_B T_0 (m_1 \omega_1^2)^{-1}}{\Sigma^4} - \frac{2r(q_1 - q_{10})q_2}{\Sigma^2} \right\} \right\} \tag{44}$$

which is nothing but the probability distribution of the 2*D* correlated model (11) by means of the identifications

$$(x, y) \longleftrightarrow (q_1, q_2)$$
$$\mu_x \longleftrightarrow q_{10}$$
$$\sigma \longleftrightarrow \sqrt{k_B T_0 (m_1 \omega_1^2)^{-1}} \tag{45}$$
$$(1 - r^2)^{-1} \longleftrightarrow \beta k_B T_0 = \frac{T_0}{T}.$$

From the last line of (45) one obtains

$$1 - r^2 = \frac{T}{T_0}. \tag{46}$$

Let us show that this equation is a particular case of the formula (40): since the determinant of the covariance matrix is $\Sigma^4 (1 - r^2)$ and the determinant of the Hessian of the potential energy evaluated at (q_{10}, q_{20}) is $(1 - r^2) m_1 m_2 \omega_1^2 \omega_2^2$ then one can replace both expressions in (40) with $n = 1$, thus obtaining

$$\Sigma^4 (1 - r^2)^2 m_1 m_2 \omega_1^2 \omega_2^2 = (k_B T)^2.$$

With the help of (41) one can recast this equation as

$$\Sigma^4 (1 - r^2)^2 (k_B T_0)^2 \frac{1}{\Sigma^4} = (k_B T)^2$$

from which one obtains $1 - r^2 = \frac{T}{T_0}$.

Considering the temperature T as an external parameter one can study the phase transitions of the system as T varies. Also, we set the reference temperature T_0 as the room temperature. Given that the parameter τ is arbitrary, we choose τ as

$$\tau = \frac{1}{1 - \frac{T}{T_0}}.$$

This choice for τ is convenient since one has

$$\tau \longrightarrow \infty \quad \text{if and only if } T \longrightarrow T_0. \tag{47}$$

---

[3] For instance, the room temperature ~20 °C (293.15 K).

In this way, the asymptotic limit $\tau \to \infty$ is identified with the limit $T \to T_0$, and therefore, the transition towards high temperatures can be studied by means of the limit $\tau \to \infty$. This transition express the behavior of the correlations between the oscillators when the bath temperature pass from a finite value (which is lower than T_0) to the room temperature. Physically, at the room temperature is expected that if the energy $k_B T_0$ delivered by the bath to each oscillator is larger than the energies of them (i.e., $k_B T_0 \gg m_{1,2}\omega_{1,2}^2$) then as a result of the thermal agitation the correlations between the oscillators tend to be canceled. Indeed, from (46) one can see that $r \to 0$ when $T \to T_0$.

Now let us see that this phase transition is characterized in terms of the mixing level of the IGEH. When r is vanishingly small, the following approximations hold:

$$(\sqrt{1-r^2})^{-1-\frac{1}{|r|}} \approx 1 - r^2 \left(-\frac{1}{2} - \frac{1}{2|r|} \right) = 1 + \frac{1}{2}|r| + \frac{1}{2}r^2$$

$$(1 + |r|)^{-1-\frac{1}{|r|}} \leq 1.$$

Using these approximations, and neglecting terms of order r^2, in the formula (29) of $F(p)$ one obtains that $F(p) \lesssim |r|$ holds for $|r| \ll 1$. Replacing this inequality in (28) one has that

$$|C(f_1, f_2, \tau)| \lesssim |r| \, \|f_1 f_2\|_1, \quad \text{for } |r| \ll 1.$$

In turn, since $r \to 0$ when $\tau \to \infty$ this equation implies

$$\lim_{\tau \to \infty} |C(f_1, f_2, \tau)| \lesssim \lim_{\tau \to \infty} |r| \, \|f_1 f_2\|_1 = 0. \tag{48}$$

According to the definition (25) then the system is *IG mixing*. This is the regime of null correlations corresponding to the region of the curve of $F(p)$ around at $r = 0$, as can be seen in Fig. 1. When $r = 0$ the probability distribution (44) can be factorized as the product of its marginals, and therefore, from the Eq. (26) one has that the system is *IG Bernoulli*.

Moreover, since the scalar curvature for the 2D correlated model is $R = -\frac{1}{2}(1 - r^2)$ then one obtains

$$R = -\frac{T}{2T_0}. \tag{49}$$

The formula (49) expresses the connection between the thermodynamics of the canonical ensemble and the information geometry of the system of coupled oscillators. It can be seen that the statistical model behaves as an "intermediary" between the thermodynamic parameters and the geometrical quantities of the statistical manifold. As a consequence, the covariance matrix determinant of the statistical model is determined by the Boltzmann factor, thus linking the temperature with all the geometrical quantities like the metric tensor, the scalar curvature, etc.

From Eq. (49) it follows that the scalar curvature decreases as the bath temperature increases up to reach a minimum value $R = -\frac{1}{2}$ at the room temperature, where the correlation coefficient is zero. On the other hand, when the temperature tends to zero the scalar curvature takes a maximum value $R = 0$ which corresponds to the maximally correlated case $r \to \pm 1$. This reflects the intuitively image that in absence of thermal agitation the correlations remain present.

Finally, it should be noted that this analysis is consistent with the Cafaro's criteria of global chaos: as the temperature grows the dynamics become more chaotic and the scalar curvature turns out more negative. From the point of view of the IGEH this is characterized in terms of the mixing level, in which the breaking up of the coupling between the oscillators at the room temperature is expressed by means of the cancellation of the IG correlation in the asymptotic limit.

7.2. 2 × 2 Gaussian Orthogonal Ensembles (GOE)

In Gaussian Orthogonal Ensembles theory one deals with the probability distribution $p(H_{11}, H_{12}, \ldots, H_{nn})$ for the Hamiltonian matrix elements assuming that the H_{ij} are uncorrelated [20,21]. Then in the framework of information geometry one could try to describe them by defining a microspace $x_1, x_2, \ldots, x_n$ and a macrospace $\theta_1, \theta_2, \ldots, \theta_m$ in a suitable way.

In order to characterize the GOE within a statistical model we study a correlated ensemble of 2×2 matrices. We take the microspace as the Hamiltonian matrix elements $\{H_{11}, H_{22}, H_{12}, H_{21}\}$ and define the macrospace as follows. For the sake of simplicity, we choose the macrospace in such way that only H_{11} and H_{22} are correlated, and that the mean values of all variables are zero, except for the mean value corresponding to H_{11} which is equal to μ. Also, we consider that the variance of H_{11}, H_{12} and H_{21} are the same, denoted by σ. Moreover, in order to study how independent the diagonal Hamiltonian elements are, we restrict the dynamics by considering that $r \in [-1, 1]$ is the correlation coefficient between H_{11} and H_{22}, and that the product of the covariances between H_{11} and H_{22} is a constant Σ^2. Taking this into account, the resulting macrospace is $\{(\mu, \sigma) \in \mathbb{R} \times (0, \infty)\}$ and the correlated probability distribution is given by[4]

$$p(H_{11}, H_{22}, H_{12}, H_{21}; \mu, \sigma, r)$$

$$= \frac{1}{2\pi \Sigma^2 \sqrt{1-r^2}} \exp\left(-\frac{1}{2(1-r^2)} \left[\frac{(H_{11} - \mu)^2}{\sigma^2} + \frac{H_{22}^2 \sigma^2}{\Sigma^4} - \frac{(2r(H_{11}-\mu)H_{22})}{\Sigma^2} \right] \right) \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(H_{12}^2 + H_{21}^2)}{2\sigma^2} \right)$$

---

[4] Note that, since the GOE correspond to the orthogonal class of Hamiltonians then one has that $H_{12} = H_{21}$. However, in the formalism of Random matrices and for the orthogonal case, the volume element $dH_{11}dH_{22}dH_{12}dH_{21}$ (as if H_{12} and H_{21} were independent variables) is the real Lebesgue measure of $\mathbb{R}^4$ and must to be taken into account in order to normalize the probability distribution [20].
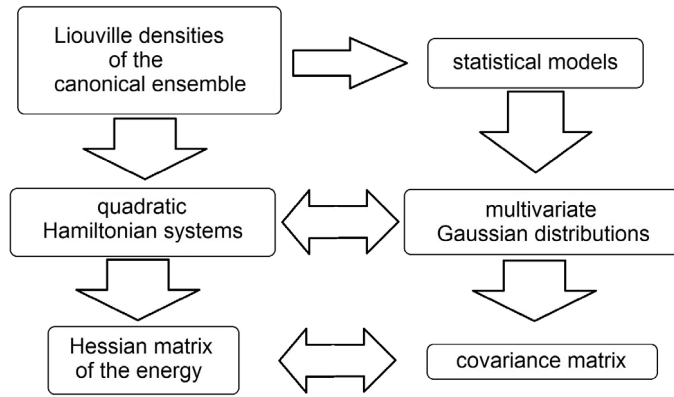
Fig. 2. The Liouville density of the canonical ensemble defines a family of PDFs parameterized by macroscopic quantities which constitutes a statistical model. For the special case of the quadratic Hamiltonians, the correspondence between the Liouville densities and the multivariate Gaussian distributions is one-to-one, and therefore the same happens between the Hessians energy and the covariance matrices respectively.

where the correlation coefficient r is considered as an external parameter and since the correlation between H_{11} and H_{22} is in terms of r then Σ can be taken as a fixed constant. In turn, given that H_{11} and H_{22} are the only microvariables correlated to each other and since the transitions of the dynamics depend fundamentally on the correlations, then one can reasonably neglect H_{12} and H_{21}. Analogously as it was made in the pair of oscillators, one can integrate the correlated probability distribution over H_{12} and H_{21}, thus obtaining

$$p(H_{11}, H_{22}; \mu, \sigma, r) = \frac{1}{2\pi \Sigma^2 \sqrt{1 - r^2}} \exp\left(-\frac{1}{2(1 - r^2)}\left[\frac{(H_{11} - \mu)^2}{\sigma^2} + \frac{H_{22}^2 \sigma^2}{\Sigma^4} - \frac{(2r(H_{11} - \mu)H_{22})}{\Sigma^2}\right]\right) \qquad (50)$$

which is nothing but the 2D correlated model, i.e. (11) and (50) identical by renaming H_{11} and H_{22} as x and y respectively. Then, the non vanishing components of the Ricci tensor R_{ij} and the Ricci scalar curvature R are given by the Eqs. (14) and (15)

$$R = g^{11}R_{11} + g^{22}R_{22} = -\frac{1}{2}(1 - r^2), \quad \text{with } R_{11} = -\frac{1}{4\sigma^2}, \ R_{22} = -\frac{1}{\sigma^2}.$$

Three remarks follow. First, the statistical manifold has a curvature which is negative for all values of the correlation coefficient $r \in [-1, 1]$. Based on the Cafaro's criterium above, this simply means that the dynamics in macrospace (μ, σ) is chaotic for all r.

Second, the 2×2 GOE case corresponds to $r = 0$ and $\Sigma = \sigma$, thus having the minimum value of the scalar curvature

$$R_{GOE} = R(r = 0) = -\frac{1}{2} = R_{\min} \quad \text{(GOE, most chaotic case)}.$$

In this case the correlated probability distribution is the product of their marginals and thus, the model is *IG Bernoulli*. Therefore, one can see that the GOE corresponds to the strongest level of the IGEH and this can be considered as a characterization of the Gaussian ensembles from an information geometric point of view.

Third, for the strongly correlated case that corresponds to $|r| \sim 1$ one has

$$R(|r| \to 1) = 0 \quad \text{(strongly correlated case)}$$

which can be interpreted, by the Cafaro's criterium of global chaos, as the case when the dynamics is the least chaotic of all.

7.3. A panoramic outlook of the IGEH and of the canonical ensemble in curved statistical models

Here we summarize the aspects of our proposal that can provide innovative tools within the context of the information geometry for characterizing the dynamics of a system in terms of the chosen statistical model. For this, below we provide two schematic diagrams, Figs. 2 and 3, showing the content of the two main contributions of this work and its physical relevance from a panoramic outlook.

8. Conclusions

We have proposed an extension of the ergodic, mixing and Bernoulli levels in the context of information geometry, that we called *information geometric ergodic hierarchy (IGEH)*, and we applied it to characterize: (i) the phase transitions of a pair of interacting harmonic oscillators in the canonical ensemble and (ii) the 2×2 Gaussian Orthogonal Ensembles. The relevance and novelty of our main contributions, i.e. the IGEH and the information geometric characterization of the Liouville densities of the canonical ensemble expressed by the Eqs. (22)–(26) and (31)–(40) respectively, lie in the following remarks:
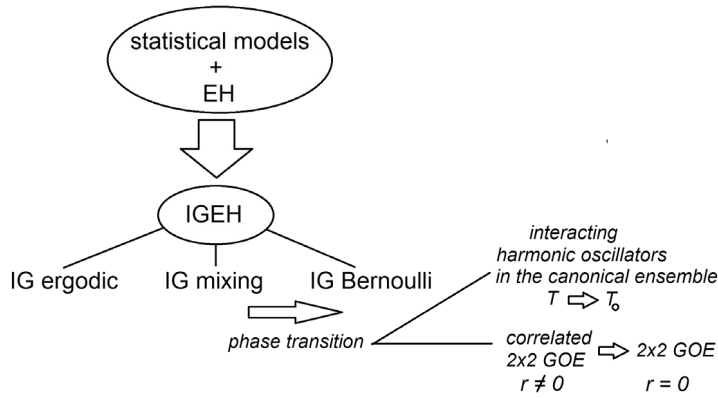
Fig. 3. A diagram showing the way in which the IGEH can characterize phase transitions in terms of the passage from one of their levels to other, and in different applications: a pair of interacting harmonic oscillators in the canonical ensemble and the 2×2 GOE.

- Statistical models provides a unified scenario for approaches involving correlations between microscopic variables. This was illustrated with a correlated 2×2 GOE by adding correlations between two variables and showing that, this modification attenuates the chaotic dynamics on macrospace by increasing the scalar curvature, accordingly to the Cafaro's criterium of global chaos.
- The IGEH generalizes the chaos characterization of the ergodic hierarchy by quantifying the statistical independence between the microvariables (instead of subsets of phase space) of the statistical model. This is performed in the asymptotic limit of large values of the time-like parameter which is expressed in terms of upper bounds of the IG correlation as the measure $F(p)$ for the case of the $2D$ correlated model.
- The association between multivariate Gaussian distributions and quadratic Hamiltonians can be useful for studying the type of stability that present the dynamics in their equilibrium points in the context of the information geometry.
- Geometrical notions and the Cafaro's criterium of global chaos can be related with the levels of the IGEH. The 2×2 GOE case belonging to the most chaotic level, the IG Bernoulli, has an associated minimum negative value of the scalar curvature $R_{GOE} = -\frac{1}{2}$.
- By obtaining upper bounds $F(p)$ on the IG correlation for a specific family of probability distributions, as exemplified by the curve of Fig. 1, one could study geometrical phase transitions moving along curves $F(p)$ as an external parameter r is varied.

Acknowledgments

Appendix. Distinguishability measure $F(p)$ (formula (29))

Proof. Replacing (11) in the definition of $F(p)$ one obtains

$$F(p) = \max_{\{(x,y) \in \mathbb{R}^2\}} \left| \frac{1}{2\pi \Sigma^2 \sqrt{1-r^2}} \exp\left(\frac{-1}{2(1-r^2)} \left[\frac{(x-\mu_x)^2}{\sigma^2} + \frac{y^2\sigma^2}{\Sigma^4} - \frac{2r(x-\mu_x)y}{\Sigma^2} \right] \right) \right.$$
$$\left. - \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu_x)^2}{2\sigma^2} \right) \frac{\sigma}{\sqrt{2\pi\Sigma^4}} \exp\left(-\frac{y^2\sigma^2}{2\Sigma^4} \right) \right|.$$

Now, by defining the following adimensional variables $\widetilde{x} = \frac{x-\mu_x}{\sigma}$ and $\widetilde{y} = y\frac{\sigma}{\Sigma^2}$ one can recast $F(p)$ as

$$F(p) = \frac{1}{2\pi \Sigma^2} \text{máx}_{\{(\widetilde{x},\widetilde{y}) \in \mathbb{R}^2\}} \left| \frac{1}{\sqrt{1-r^2}} \exp\left(\frac{-1}{2(1-r^2)} \left(\widetilde{x}^2 + \widetilde{y}^2 - 2r\widetilde{x}\widetilde{y} \right) \right) - \exp\left(-\frac{\widetilde{x}^2 + \widetilde{y}^2}{2} \right) \right|. \tag{A.1}$$

Therefore, in order to calculate $F(p)$, it is enough to study the maximum and minimum on $\mathbb{R}^2$ of the function $G_r(x,y) = \frac{1}{\sqrt{1-r^2}} \exp\left(\frac{-1}{2(1-r^2)} \left(x^2 + y^2 - 2rxy \right) \right) - \exp\left(-\frac{x^2+y^2}{2} \right)$ for each value of the parameter $r \in [-1, 1]$. For this, it is convenient to write $G_r(x,y) = g_1(x,y) - g_2(x,y)$ with $g_1(x,y) = \frac{1}{\sqrt{1-r^2}} \exp\left(\frac{-1}{2(1-r^2)} \left(x^2 + y^2 - 2rxy \right) \right)$ and $g_2(x,y) = \exp\left(-\frac{x^2+y^2}{2} \right)$.

Then, one must to find the critical points of $G_r(x, y)$ from the equations

$$\frac{\partial G_r(x, y)}{\partial x} = \frac{\partial g_1(x, y)}{\partial x} - \frac{\partial g_2(x, y)}{\partial x} = g_1(x, y) \left(\frac{-1(x - ry)}{(1 - r^2)} \right) - g_2(x, y)(-x) = 0 \tag{A.2}$$

$$\frac{\partial G_r(x, y)}{\partial y} = \frac{\partial g_1(x, y)}{\partial y} - \frac{\partial g_2(x, y)}{\partial y} = g_1(x, y) \left(\frac{-1(y - rx)}{(1 - r^2)} \right) - g_2(x, y)(-y) = 0.$$

From (A.2) it follows that $y \frac{\partial G_r(x,y)}{\partial x} - x \frac{\partial G_r(x,y)}{\partial y} = 0$, thus

$$y g_1(x, y) \left(\frac{-1(x - ry)}{(1 - r^2)} \right) + g_2(x, y)xy - x g_1(x, y) \left(\frac{-1(y - rx)}{(1 - r^2)} \right) - g_2(x, y)xy = 0.$$

Since $g_1(x, y)$ is an exponential then is always positive. Then one obtains

$$g_1(x, y) \left(\frac{-y(x - ry) + x(y - rx)}{(1 - r^2)} \right) = 0 \implies g_1(x, y) \left(\frac{r(y^2 - x^2)}{(1 - r^2)} \right) = 0 \implies r = 0 \quad \text{or} \quad y = \pm x.$$

If $r = 0$ then it is clear that p is equal to the product of their marginals and it follows that the infinite norm is zero. Assuming $r \neq 0$ and replacing $y = \pm x$ in the formula of $G_r(x, y)$ then one obtains a function that only depends on x, that we can denote as $\varphi_r(x)$. Then one has that

$$\varphi_r(x) = \frac{1}{\sqrt{1 - r^2}} \exp\left(\frac{-x^2}{(1 - r^2)} (1 \mp r) \right) - \exp\left(-x^2 \right). \tag{A.3}$$

Thus, the value of $F(p)$ is given by the maximum value of $|\varphi_r(x)|$ with $x \in \mathbb{R}$. For this, we calculate the critical points of $\varphi_r(x)$ by making its derivative equal to zero, thus obtaining

$$\begin{aligned}\frac{d\varphi_r(x)}{dx} &= \frac{1}{\sqrt{1 - r^2}} \exp\left(\frac{-x^2}{(1 - r^2)} (1 \mp r) \right) \left(\frac{-2x}{(1 - r^2)} (1 \mp r) \right) + 2x \exp\left(-x^2 \right) \\ &= 2x \left[\exp\left(-x^2 \right) - \frac{1}{\sqrt{1 - r^2}} \exp\left(\frac{-x^2}{(1 - r^2)} (1 \mp r) \right) \left(\frac{1}{(1 - r^2)} (1 \mp r) \right) \right] = 0 \\ &\implies x = 0 \quad \text{or} \quad \exp\left(-x^2 \right) - \frac{1}{\sqrt{1 - r^2}} \exp\left(\frac{-x^2}{(1 - r^2)} (1 \mp r) \right) \left(\frac{1}{(1 - r^2)} (1 \mp r) \right) = 0.\end{aligned}$$

So the critical points of $\varphi_r(x)$ satisfy

$$x = 0 \quad \text{or} \quad \frac{1}{\sqrt{1 - r^2}} \exp\left(\frac{-x^2}{(1 - r^2)} (1 \mp r) \right) \left(\frac{1}{(1 - r^2)} (1 \mp r) \right) = \exp\left(-x^2 \right). \tag{A.4}$$

Replacing (A.4) in (A.3) one has the value of $\varphi_r(x)$ in its critical points x_c

$$\varphi_r(x_c) = \left\{ \varphi(0) = \frac{1}{\sqrt{1 - r^2}} - 1, \ \pm r \exp\left(-x_c^2 \right) \right\}. \tag{A.5}$$

In order to find explicitly $\exp\left(-x_c^2 \right)$ one must solve the second relation of the Eq. (A.4), taking the natural logarithm in both sides of it then one has

$$-\frac{x_c^2(1 \mp r)}{(1 - r^2)} + \ln\left(\frac{1}{\sqrt{1 - r^2}(1 - r^2)} (1 \mp r) \right) = -x_c^2.$$

That is,

$$\exp\left(-x_c^2 \right) = \left(\frac{1}{\sqrt{1 - r^2}(1 \pm r)} \right)^{\left(\frac{1 \pm r}{\mp r} \right)}. \tag{A.6}$$

Then, from (A.6) and (A.5) one obtains the following values of $\varphi_r(x_c)$ in its critical points

$$r \left(\sqrt{1 - r^2}(1 + r) \right)^{-1 - \frac{1}{r}}, \qquad -r \left(\sqrt{1 - r^2}(1 - r) \right)^{-1 + \frac{1}{r}}, \qquad \frac{1}{\sqrt{1 - r^2}} - 1. \tag{A.7}$$

Now, given that the maximum and minimum values of $G(x, y)$ are precisely the same of $\varphi_r(x)$ subject to the restriction $y = \pm x$ and since the maximum value of $|G(x, y)|$ is equal to $F(p)$, then from (A.7) one deduces that $F(p)$ is

$$F(p) = \max \left\{ |r| \left(\sqrt{1 - r^2}(1 + r) \right)^{-1 - \frac{1}{r}}, |r| \left(\sqrt{1 - r^2}(1 - r) \right)^{-1 + \frac{1}{r}}, \frac{1}{\sqrt{1 - r^2}} - 1 \right\} \tag{A.8}$$

from which follows the desired result. $\square$

References

[1] F. Weinhold, J. Chem. Phys. 65 (1976) 559. 63, 2479, 2488 (1975).
[2] G. Ruppeiner, Phys. Rev. A 20 (1979) 1608; Rev. Modern Phys. 67 (1995) 605. Erratum: 68, 313 (1996).
[3] C.R. Rao, Differential metrics in probability spaces, in: Differential Geometry in Statistical Inference, Institute of Mathematical Statistics, Hayward, CA, 1987, (Chapter);
C.R. Rao, Bull. Calcutta Math. Soc. 37 (1945) 81.
[4] S. Amari, H. Nagaoka, Methods of Information Geometry, Oxford University Press, Oxford, UK, 2000.
[5] R.S. Ingarden, Tensor (N.S.) 30 (1976) 201.
[6] H. Janyszek, Rep. Math. Phys. 24 (1986) 1. 11.
[7] S. Abe, Phys. Rev. E 68 (2003) 031101.
[8] J. Naudts, Open Syst. Inf. Dyn. 12 (2005) 13.
[9] M. Portesi, A. Plastino, F. Pennini, Physica A 365 (2006) 173–176; Physica A 373 (2007) 273–282.
[10] C. Tsallis, J. Stat. Phys. 52 (1988) 479.
[11] D. Bures, Trans. Amer. Math. Soc. 135 (1969) 199.
[12] G. Verdoolaege, AIP Conf. Proc. 1641 (2014) 564–571; Rev. Sci. Instrum. 85 (2014) 11E810.
[13] C. Cafaro, Chaos Solitons Fractals 41 (2009) 886–891.
[14] C. Cafaro, S. Mancini, Physica D 240 (2011) 607–618.
[15] C. Cafaro, A. Giffin, C. Lupo, S. Mancini, Open Syst. Inf. Dyn. 19 (2012) 1250001.
[16] A. Giffin, S.A. Ali, C. Cafaro, Entropy 15 (2013) 4622–4623.
[17] A.J. Lichtenberg, M.A. Lieberman, Regular and Chaotic Dynamics, in: Applied Mathematical Sciences, Springer, Berlin, 2010.
[18] J. Berkovitz, R. Frigg, F. Kronz, Stud. Hist. Philos. Modern Phys. 37 (2006) 661–691.
[19] A. Lasota, M. Mackey, Probabilistic Properties of Deterministic Systems, Cambridge Univ. Press, Cambridge, 1985.
[20] H. Stockmann, Quantum Chaos - An Introduction, Cambridge Univ. Press, Cambridge, 1999.
[21] F. Haake, Quantum Signatures of Chaos, Springer-Verlag, Heidelberg, 2001.
[22] G. Casati, B. Chirikov, Quantum Chaos: Between Order and Disorder, Cambridge Univ. Press, Cambridge, 1995.
[23] G. Casati, T. Prosen, Phys. Lett. A 72 (2005) 032111.
[24] O. Bohigas, M. Giannoni, C. Schmit, Phys. Rev. Lett. 52 (1984) 1.
[25] I. Gomez, M. Castagnino, Physica A 393 (2014) 112–131.
[26] I. Gomez, M. Castagnino, Chaos Solitons Fractals 68 (2014) 98–113.
[27] I. Gomez, M. Castagnino, Chaos Solitons Fractals 70 (2015) 99–116.
[28] I. Gomez, M. Losada, S. Fortin, M. Castagnino, M. Portesi, Internat. J. Theoret. Phys. 7 (2015) 2192–2203.
[29] I.S. Gomez, M. Portesi, arXiv:1503.02751v3 [quant-ph] (2017).
[30] O.E. Barndorff-Nielsen, Information and Exponential Families in Statistical Theory, J. Wiley and Sons, New York, 1978.
[31] J. Naudts, J. Ineq. Pure Appl. Math. 5 (2004) 102.
[32] P.D. Grünwald, A.P. Dawid, Ann. Statist. 32 (2004) 1367–1433.
[33] A. Ohara, T. Wada, J. Phys. A 43 (2010).
[34] J. Naudts, Entropy 10 (2008) 131–149.
[35] S. Amari, A. Ohara, H. Matsuzoe, Physica A 391 (2012) 4308–4319.
[36] M. Molitor, J. Geom. Phys. 70 (2013) 54–80.
[37] H. Matsuzoe, Differential Geom. Appl. 35 (2014) 323–333.
[38] F. Zang, Y. Shi, R. Wang, Physica A 468 (2017) 552–565.