Complex Systems are ubiquitous in nature and man-made systems. In natural sciences, in social and economic models and in mathematical constructions are studied and analyzed, are applied in practical problems but without a clear and universal definition of "complexity", let alone classification and quantification. Following the "three-level scheme" of physical theories, observations/experiments, phenomenology, microscopic interactions, we need, starting from the experience of observation to establish appropriate phenomenological parameters and concepts, and in conjunction with a possible knowledge of the nature of microscopic structures to deepen our understanding of a particular system which we "understand as complex". Information Geometry seems to be a useful phenomenological framework, which using generalized entropies, provides some classification and quantification tools. But we need the next level, microscopic structure and interactions of the parts of complex systems. A useful direction is the conceptual niche of hyper-networks and super graphs, where a strong involvement of algebra offers concrete techniques. We believe that appropriate algebraic structures may systematize our approach to microscopic structures of complex systems, and help associate the information geometric phenomenology with concrete properties. In this paper after a short discussion of the problem of "definition of complexity", we introduce our information geometric quantities derived from generalized entropies. Then we present our results of application of information geometry for classification of complex systems. Finally we present our ideas for an abstract algebraic approach which may offer a framework for the microscopic study of complex systems.

AMS Subject Classification: 94A17
Keywords: complex systems, generalized entropies, information geometry
DOI: https://doi.org/10.33581/1561-4085-2020-23-2-212-220

1. Defining, quantifying, and classifying complex systems

Some general characteristics of a complex system [1] are

- many heterogeneous parts
- complicated transition laws
- unexpected or unpredictable emergence
- sensitive dependence on initial conditions
- path-dependent dynamics
- network connectivities
- multiple subsystem dependencies
- dynamics emerge from interactions of autonomous agents
- self-organization into new structures and behaviors
- non-equilibrium and far-from equilibrium dynamics
- discrete dynamics with combinatorial explosion
- adaptation to changing environments
- co-evolving subsystems
- ill-defined boundaries
- multilevel dynamics
- feedback
- confounding behavior
- unrepeatable experiments
- misleading data
globality (everything affects everything else)

This is a list of answers that most people would give to the question "What a Complex System is". These are cases of systems that we consider as complex for one or another reason. We feel that complex is something more or different from complicated and the problem is how to define quantitatively the complexity and further to classify complexity in various classes. There are three general approaches to the definition and classification of Complex Systems.

**Case study approach.** There are systems which we attribute the property of being complex and we construct quantitative tools to analyze and classify them in categories in partially ordered sets. [7, 8, 26, 29, 32, 38–42, 45, 46] Examples are:
- Theory of Dynamical Systems,
- Neuroscience,
- Financial Markets.

Though the conceptual tools that are introduced are very effective they may not cover all cases of systems we "consider" as complex.

**Axiomatic/Abstract approach.** The aim of this approach is to establish universal properties which cover most systems considered as complex. This is objective and rigorous but in general the quantification is non-computable [20, 22, 37].

**Analysis of Entropies and derived concepts.** This approach belongs mainly to the phenomenological level, is based on general observations and may have constructs emanating from the microscopic level [9, 12, 19, 21, 25, 30, 33–35]. An offspring of statistical behavior is the Information Geometry.

2. Generalized entropies and complex systems

The most famous and well established "three level scheme" in physics is:

- thermal phenomena are organized observations,
- thermodynamics is Phenomenology,
- statistical mechanics is Microscopic Theory,

Entropy is one of the most fundamental phenomenological concepts that transcends physics and plays a key role in Information Theory.

2.1. Entropic functionals in Information Theory

Assuming the four Shannon–Khinchin Axioms we have a unique entropy functional, the Boltzmann-Gibbs Entropy

\[ S[p] = - \sum_j p(j) \ln p(j), \quad \sum_j p(j) = 1. \]  

This covers many classical cases with extreme success. But there exist situations where it seems more appropriate or more useful to introduce other entropy functionals capable of giving more specialized information on some systems. Such generalized functionals are Renyi's entropy

\[ S_q[p] = \frac{1}{1-q} \ln \left( \sum_j p(j)^q \right) \]  

and Tsallis entropy

\[ S_{Tsallis}^q[p] = \frac{1}{1-q} \left( \sum_j p(j)^q - 1 \right). \]

It is interesting and very important that these entropic functionals depend on a parameter, which may be related to specific system properties. I would like to point out that these are phenomenological parameters, as is the friction coefficient of elementary physics. There have been similar constructions of Information Geometry, based on various generalizations of the entropy [2–4, 6, 10, 18, 23, 24, 27, 28, 31, 36].

2.2. A two-parameter entropic functional

Hanel and Thurner [13–17] in their quest for classification of complex systems derived a two-parameter entropy functional. Their result is that complex systems are classified in five broad classes. Their generalization came from the
assumption that the functional does not satisfy additivity. This gave a form that depends on two parameters \((c, d)\) with \((c, d) = (1, 1)\), and \((c, d) = (c, 0)\) corresponding to Boltzmann–Gibbs and Tsallis entropies respectively. The Hanel–Thurnel two parameter entropy functional is

\[
S_{c,d}[p] = \frac{e \sum_{i} \Gamma(d + 1, 1 - c \ln p_i) - c}{1 - c + cd}
\]

where \(M\) is the number of potential outcomes and \(\Gamma(a, b) = \int_{b}^{\infty} dt t^{a-1} \exp(-t)\) is the incomplete Gamma-function. The Boltzmann–Gibbs entropy is recovered for \((c, d) = (1, 1)\), while for the Tsallis entropy we have \((c, d) = (c, 0)\). The associated distribution function is the generalized exponential

\[
E_{c,d,r}(x) = \exp\left( - \frac{d[W_k(B(1 - x/r)^{1-c}) - W_k(B)]}{1 - c} \right)
\]

where \(r = (1 - c + cd)^{-1}\) and \(B = \frac{(1-c)^r}{1-(1-c)^r} \exp(\frac{(1-c)r}{1-(1-c)^r})\). The function \(W_k\) is the \(k\)-th branch of the Lambert \(W\)-function which is a solution of the equation \(x = W(x) \exp(W(x))\). This generalized exponential is the inverse function of the generalized logarithm:

\[
\Lambda_{c,d,r}(x) = r - rx^{c-1}\left[1 - \frac{1 - (1-c)r}{rd} \ln x\right]^d.
\]

3. Information geometry and classification of complex systems

3.1. Geometry from probability distributions

Here we present only necessary concepts in order to establish the notation. We refer to the bibliography for the details [2–5]. Let

\[
S = \{p_\xi = p(x; \xi) | \xi = [\xi^1, \ldots, \xi^n] \in \Xi\}
\]

be a parametric family of probability distributions on \(\mathcal{X}\). This is a \(n\)-dimensional parametric statistical model. Given the \(N\) observations \(x_1, \ldots, x_N\) the classical estimation problem concerns the statistical methods that may be used to detect the true distribution, that is to estimate the parameters \(\xi\). To this purpose, an appropriate estimator is used for each parameter. These estimators are maps from the space of random variables to the parameter space of the model. The quality of estimation is measured by the variance–covariance matrix \(V_\xi = [v^i_j]_\xi\) where

\[
v^i_j_\xi = E_\xi[(\hat{\xi}^i(X) - \xi^i)(\hat{\xi}^j(X) - \xi^j)].
\]

Suppose that the estimators are unbiased, namely

\[
E_\xi[\hat{\xi}(X)] = \xi, \quad \forall \xi \in \Xi.
\]

Then a lower bound for the estimation error is given by the Cramer-Rao inequality

\[
V_\xi(\hat{\xi}) \geq G(\xi)^{-1}
\]

where \(G(\xi) = [g_{ij}(\xi)]\).

\[
g_{ij}(\xi) = E_\xi[\partial_i l(x; \xi) \partial_j l(x; \xi)]
\]

is the classical Fisher matrix,

\[
l_\xi = l(x; \xi) = \ln p(x; \xi)
\]

is the score function. As it has been shown the Fisher matrix provides a metric on the manifold of classical probability distributions. This metric, according to the theorem of Cencov [11], is the unique metric which is monotone under the transformations of the statistical model. This means that if the map \(F : \mathcal{X} \rightarrow \mathcal{Y}\) induces a model \(S_F = \{q(y; \xi)\} \) on \(\mathcal{Y}\) then

\[
G_F(\xi) \leq G(\xi).
\]

That is, the distance of the transformed distributions is smaller than the original distributions. Thus the monotonicity of the metric is intuitively related to the fact that in general we loose distinguishability of the distributions from any transformation of the information. The metric defined in this way is the ordinary Fisher metric. Using the Levi–Civita connection the corresponding Riemannian structure is constructed. In this geometry the scalar curvature is a quantification of the information manifolds. But there is a further development connected with the existence of connections different from Levi–Civita. These are certain pairs of connections satisfying a duality property with respect to the Fisher metric and playing a fundamental role in the estimation theory. An important case is the dually flat connections.

Нелинейные явления в сложных системах Т. 23, № 2, 2020

Demetris P. K. Ghikas
3.2. Geometry from divergences

A further extension is the derivation of the differential structure from relative entropies or divergence functions. These are quasi-distances and particular cases have been used with various names. Let \( p, q \) be distribution functions considered as points in an information manifold. A divergence \( D(p\|q) \) satisfies the property

\[
D(p\|q) \geq 0 \text{ and } D(p\|q) = 0 \iff p = q.
\]

Now, considering the function \( D(p\|p + dp) \) and expanding to third order we get a metric and a connection characterized by \( D \):

\[
g^{ij}_{ij} = -\partial_i \partial_j D(p\|p')|_{p'=p},
\]

\[
\Gamma^D_{ij,k} = -\partial_i \partial_j \partial_k D(p\|p')|_{p'=p}
\]

where \( \partial_i = \frac{\partial}{\partial x^i} \) and \( \partial_i' = \frac{\partial}{\partial x'^i} \). A fundamental concept of great practical usefulness in the estimation theory is the duality. Given a metric and two connections \( (g, \nabla, \nabla') \) the connections are dual with respect to the metric if

\[
\partial_k g_{ij} = \Gamma_{ki,j} + \Gamma^*_{kj,i}
\]

holds. From the geometry coming from a divergence a dual structure is obtained by defining \( \Gamma^{D^*}_{ij,k} = -\partial_k \partial_i \partial_j D(p\|p')|_{p'=p} \). There is a general family of divergences, the so called \( f \)-divergences, which are generalizations of the known Kullback–Leibler divergence. In the statistical applications a special role is played by the dually flat connections. In this case there exist dual coordinate systems on the manifold, \( [\theta^i], [\eta_i] \) and functions \( \psi \) and \( \phi \) such that

\[
\theta^i = \partial^i \phi, \quad \eta_i = \partial_i \psi, \quad g_{ij} = \partial_i \partial_j \psi, \quad g^{ij} = \partial^i \partial^j \phi.
\]

This is the Legendre transformation with the corresponding potential function \( \psi \) and \( \phi \). There is a canonical divergence which is uniquely defined for dually flat manifolds

\[
D(p\|q) \equiv \psi(p) + \phi(q) - \theta^i(p) \eta_i(q).
\]

3.3. Exponential families

Exponential families are a general family of probability distributions. The standard form is

\[
p(x, \theta) = \exp \left\{ \sum_i \theta^i x_i - \psi(\theta) \right\}.
\]

Exponential families have an inherent dually flat structure. And this offers a natural route to construct geometries for generalized exponentials which are related to generalized entropies. A discrete distribution has the form of an exponential family if we define appropriately the parameters

\[
\theta^i = \log \frac{p(x_i)}{p(x_0)} = \log \frac{\xi^i}{1 - \sum_{j=1}^n \xi^j},
\]

\[
\psi(\theta) = -\log p(\theta) = -\log \left\{ 1 - \sum_{j=1}^n \xi^j \right\}
\]

3.4. The \((c, d)\)-exponential family [43]

Starting from the \((c, d)\)-generalized logarithm we get the parameters

\[
\theta^i = r p_0^{c-1} \left[ 1 - \frac{1 - (1 - c)r}{dr} \ln p_0 \right]^d,
\]

\[
-x_i = \delta_i(x) = \begin{cases} x_i & i = 1, \ldots, n; \\ 0 & x \neq x_i \\ \end{cases}
\]

\[
\psi(\theta) = -A_{c,d,r}(p_0).
\]
3.5. The \((c,d)\)-information metric

Defining the functions

\[
\Delta(x) = \frac{\mathcal{E}_{c,d,r}(x) \ W(B(1-x/r)^{1/d})}{r(1-c)\ W(B(1-x/r)^{1/d})} \times \left(1 - \frac{x}{r}\right)^{-1},
\]

(26)

\[
K(x) = x^{\frac{c-1}{d}} \left(1 - \frac{1 - (1-c)r}{rd} \ln x\right),
\]

(27)

\[
h(p) = \int \Delta(\Lambda(p(x, \theta))) \, dx
\]

\[
eq \int \Delta(x) \theta_i - \psi(\theta) \, dx
\]

(28)

we get for the canonical divergence

\[
D_{c,d,r}(p(x, \theta_1), p(x, \theta_2)) = \psi(\theta_2) - \psi(\theta_1) - [\partial_i \psi(\theta_1)](\theta_2^i - \theta_1^i),
\]

(29)

\[
D_{c,d,r}(p, q) = \frac{1}{(1-c)h(p)} \sum_{i=0}^{n} p_i W(BK(p_i)) \left(K^{-d}(p_i)K^d(q_i) - 1\right).
\]

(30)

the information metric

\[
g_{ij}(p) = \frac{\partial^2}{\partial q_i \partial q_j} D_{c,d,r}(p, q) \mid_{q=p},
\]

(31)

\[
g_{ij}(p) = \frac{1}{(1-c)h(p)} (H(p_0) + \delta_{ij} H(p_j))
\]

(32)

where

\[
H(x) = \frac{x \ W(BK(x))}{1 + W(BK(x))} \times \left(d(d-1)K^{-2}(x)[K'(x)]^2 + dK^{-1}(x)K''(x)\right).
\]

(33)

3.6. Cramer–Rao inequalities for complex systems

**Theorem** [27, 28]

Let two families of probability distributions \(P_\theta = P_\theta(x, \theta)\) and \(p_\theta = p_\theta(x, \theta)\) and the corresponding expectations \(E_\theta\) for \(F_\theta\) and \(E_p\). Let \(c\) be an estimator of \(p_\theta = p_\theta(x, \theta)\) with scale function \(F\), that is \(E_p[\xi_k] = \frac{1}{\partial F} F(\theta)\). Assume the regularity condition

Table 1. \(l_{mean}\) is the average value of the difference between the Cramer–Rao bound for the Fisher metric and that of the \((c,d)\)-metric. The average is over the values of the parameters that are not fixed but belong to given intervals. The table is meant to give a more quantitative differentiation between the 5 classes

\[
\begin{array}{ccc}
\text{class} & c & d & l_{mean} \\
\hline
1 & 1 & 1 & 0 \\
2 & 0 < c < 1 & 0 & 0.0000 \\
3 & 1 & d > 0 & -0.0184 \\
4 & 1 & 1 & 0.0465 \\
5 & 0 < c < 1 & 1 & 0.0012 \\
\end{array}
\]

holds. Let the metric \(\tilde{g}_{ij}(\theta)\) introduced above. Then for all \(u,v\) in \(\mathbb{R}^n\)

\[
\frac{u^k u^l [F_\theta[c_k c_l] - F_\theta[c_k] F_\theta[c_l]]}{u^k u^l \partial^2 \partial c_k \partial c_l F(\theta)} \geq \frac{1}{u^k u^l \tilde{g}_{ij}(\theta)}
\]

(35)

This is a generalization of the Cramer–Rao inequality. For a first indication of the dependence of the generalized Cramer–Rao bound on the parameters of various complexity classes we compare this bound with the corresponding classical one given by the Fisher metric, for the simplest case of \(n = 1\):

\[
f(c, d, p_1) = \frac{1}{N(p) g(p)} = \frac{1}{H(p_0)} \left(\frac{1}{H(p_0) + \frac{1}{p_0}}\right) \left(H(p_0) + H(p_1)\right).
\]

(36)

The corresponding function for the Fisher metric is

\[
\tilde{f}_{Fisher}(p_1) = \frac{1}{g_{Fisher}} = \frac{1}{p_0 + \frac{1}{p_1}}.
\]

(37)

We make the Hanel–Thurner choice for \(r\), namely:

\[
r = \frac{1}{1 - c + cd}.
\]

(38)
We plot \( f \) for various classes. We also have Table 1 for

\[
I_{\text{mean}} = \text{mean}_{p_1, c, d} \{ f(p_1) - f_{\text{Fisher}}(p_1) \}
\]  (39)

where "mean" means average over the corresponding non-constant parameters \( c, d \).

In the Table caption it is explained over which parameter is the average. In Figure 1 we show how the Cramer–Rao bound differs from the one given by the Fisher metric for classes 4 and 5 of complex systems. We note here that for practical reasons, for drawing the diagrams, in the corresponding cases the values of \( c \) and \( d \) were approximately 0 and approximately 1.

3.7. Scalar curvature characterization of complex systems

In Figure 2 we show the scalar curvature over corresponding Fisher metric for the classes 4 and 5.

3.8. The \((c, d)\)-manifold [44]

Now we consider \((c, d)\) as coordinates of the manifold. We have thus a two dimensional information geometry with two parameter probability distributions

\[
p(c, d, x) = -\frac{c(1 + c(d - 1)) \exp \left( d \left( W(L) + \frac{W(L(1+(c-1-cd)x)^{1/d})}{c-1} \right) \right)}{(c - 1) \exp \left( \frac{cdW(L)}{c-1} \right) - dE_{1-d} \left( \frac{-cd}{c-1}W(L) \right)}
\]  (40)

for \( 0 < c < 1, 0 < d < 1, \) and \(-\infty < x \leq 0\) where

\[
L = L(c, d) = -\left( \frac{c - 1}{cd} \right)
\]

\[
= B \left( c, \frac{1}{1 - c + cd} \right)
\]  (41)

and

\[
E_n(z) = \int_1^{+\infty} \frac{\exp(-zt)}{t^n} dt.
\]  (42)

In Figure 3 we show a graph of the area where the scalar curvature of the \((c, d)\)-manifold has minima and a curve of these minima on the \((c, d)\)-plane.
3.9. Discussion on the information geometric results

We have stressed that there is not a universally accepted definition of complexity, let alone a quantitative measure of complexity. Our proposal to use Information Geometry, derived from a two parameter entropic functional, offers a classification framework and some possibility for a quantified differentiation between different classes of complex systems. But a theoretical prediction needs to be associated with an experimental or operational verification. Information Geometry born out of the needs of Statistics offers the test of Cramer–Rao inequalities. As it is evident from the Figure 1 the inequality bounds reveal clear differentiation between the complexity classes. We expect that this may be tested practically. On the other hand, the scalar curvature differentiation of the classes is not obvious how to be tested empirically. In general in Information Geometry the statistical usefulness of the curvature is not well understood. One may hope that such a test may be constructed. These comments are related to the case of Information Geometry based on manifolds which are parameterized by $c$ and $d$ parameters of the Hanel–Thurnel entropy. It turned out that using these parameters as coordinates of the statistical manifold we obtained a very peculiar picture of the related curvature. The minima of the scalar curvature give a curve of $c$ versus $d$ on the plane $(c,d)$. What are the properties of complex systems with parameters on this curve? Our conjecture is that there exist systems with such parameters and these systems have some special properties. How can we test this conjecture? Are there statistical tests for that? We do not know. What we may propose, at this stage, is a search for a microscopic derivation of parameterized entropic functionals. In this way, fundamental properties of the systems may give rise to families of functionals with characteristic parameters. And this association may hopefully be tested operationally.
4. Conclusion: Towards an abstract algebraic framework for complex systems

So, why algebraic framework, how a complex systems may be studied in terms of an algebraic structure, what we gain from such an approach, are there algebraic exploitations for similar abstract concepts? There are three "similar abstract concepts" (beauty, chaos, complexity) where algebras offer tools coming from concrete applications. Everybody can say if something is beautiful, but beauty cannot be quantified let alone ordered. In the everyday life it is easy to consider if something is chaotic and complex, but what is the meaning of "more chaotic" and "more complex"?

The algebraic formulations can point to "similarities" through isomorphisms, and quantification through representation theory. Established associations are symmetry, transformation groups, algebras, representations for beauty; isomorphisms of dynamics, Cuntz–Krieger algebras, graph algebras for chaos; hypernetworks, algebraic topology, evolution algebras for complexity.

We propose, at the level of microscopic structure, to construct an algebra based on the interrelation of the parts; to analyze the universality classes of these algebras, isomorphisms and representations; given the, in general, multilevel structure of complex systems [1], to construct a hierarchy of algebras that is associated to this. But, there is an inherent problem that is related with all these abstract concepts. The meaning and content is contextual. They have different representation at different phenomenological levels, or different levels of framework. Beauty, Chaos and Complexity depend on the conceptual level of analysis. The challenge is to devise the algebraic framework which can capture this singular property. I finish with famous examples of this from the area of mathematics and painting. Tiles may be beautiful but not symmetric. Strange attractors show stochasticity but are deterministic. Botticelli's "The Birth of Venus" is a beautiful picture but so is Picasso’s "Les Demoiselles d’ Avignon". It is also beautiful at least for the historians of art.

References


