Regular Article

Information geometry of scaling expansions of non-exponentially growing configuration spaces

Jan Korbel[1,2], Rudolf Hanel[1,2], and Stefan Thurner[1,2,3,4,a]

[1] Section for Science of Complex Systems, CeMSIIS, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria
[2] Complexity Science Hub Vienna, Josefstädter Strasse 39, 1080 Vienna, Austria
[3] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
[4] IIASA, Schlossplatz 1, 2361 Laxenburg, Austria

Abstract. Many stochastic complex systems are characterized by the fact that their configuration space doesn't grow exponentially as a function of the degrees of freedom. The use of scaling expansions is a natural way to measure the asymptotic growth of the configuration space volume in terms of the scaling exponents of the system. These scaling exponents can, in turn, be used to define universality classes that uniquely determine the statistics of a system. Every system belongs to one of these classes. Here we derive the information geometry of scaling expansions of sample spaces. In particular, we present the deformed logarithms and the metric in a systematic and coherent way. We observe a phase transition for the curvature. The phase transition can be well measured by the characteristic length r, corresponding to a ball with radius $2r$ having the same curvature as the statistical manifold. Increasing characteristic length with respect to size of the system is associated with sub-exponential sample space growth which is related to strongly constrained and correlated complex systems. Decreasing of the characteristic length corresponds to super-exponential sample space growth that occurs for example in systems that develop structure as they evolve. Constant curvature means exponential sample space growth that is associated with multinomial statistics, and traditional Boltzmann-Gibbs, or Shannon statistics applies. This allows us to characterize transitions between statistical manifolds corresponding to different families of probability distributions.

1 Introduction

Statistical physics of complex systems has turned into an increasingly important topic with many applications. Its main aim is to come up with a unified approach

---

[a] e-mail: stefan.thurner@meduniwien.ac.at

to understand, describe, and predict the statistical properties of a plethora of different complex systems; see, e.g., [1] for an overview. While the microscopic nature of complex systems can be very different, their statistical properties often have common features across various systems. Entropy is undoubtedly the key concept in statistical physics that connects the statistical description of microscopic dynamics with the macroscopic thermodynamic properties of a system. The notion of entropy has been adopted also in other contexts, such as information theory or statistical inference, which are concepts quite different from thermodynamics [2]. One elegant and powerful concept arising from the theory of statistical inference is that of information geometry [3,4]. It applies ideas from differential geometry to probability theory and statistics. In this context, the concept of entropy also plays a crucial role, since the metric on the statistical manifold is derived from the corresponding (relative) entropy. This, so-called Fisher-Rao metric, enables us to analyze statistical systems from a different perspective. For example, one can study critical transitions by calculating singularities of the metric [5].

In information geometry, most attention has focused on systems that are governed by Shannon entropy [3,4]. However, it is well known that many complex systems, especially strongly correlated or constrained systems, or systems with emergent components, cannot be described within the framework of Shannon entropy [1]. For this reason, a number of generalizations to Shannon entropy have been proposed; in connection with power laws [6,7], special relativity [8], multifractal thermodynamics [9], or black holes [10,11].

To classify entropies for stochastic systems of various kinds, it is natural to start with the information-theoretic foundations of Shannon entropy, i.e., the so-called Shannon-Khinchin (SK) axioms [12–14]. The first three SK axioms are usually formulated as:

– (SK1) Entropy is a continuous function of the probabilities p_i only.[1]

– (SK2) Entropy is maximal for the uniform distribution, $p_i = 1/W$.

– (SK3) Adding a state $W + 1$ to a system with $p_{W+1} = 0$ does not change the entropy of the system.

The fourth axiom is called the *composability axiom* and determines the entropy functional uniquely:

– (SK4) $H(A + B) = H(A) + H(B|A)$, where $H(B|A) = \sum p_k^A H(B|A_k)$

where $H(B|A_k)$ is the entropy of the conditional probability, $p_{B|A_k}$. In this formulation, the unique solution that is compatible with SK1-4 is Shannon entropy $H(P) = -\sum_i p_i \log p_i$. When the fourth axiom is relaxed, one can obtain a wider class of entropic functionals. First generalizations of the fourth axiom were introduced in connection with generalized additivity [15,16] group laws [17] or statistical inference [18]. These approaches are somewhat limited in scope since they all lead to a class of entropies, which can be expressed as a function of Tsallis entropy [6].

The relaxation of SK4 also naturally leads to a classification scheme of complex systems [19,20]. The main idea of this approach is to study the asymptotic scaling exponents of the entropy functional that are associated with a particular system's configuration space. These systems are associated with systems that have a sub-exponentially growing configuration space, when seen as a function of degrees of

---

[1]In several cases, entropies incorporate external parameters such as q for Tsallis entropy or c and d for (c,d)-entropies. However, these parameters are constants that characterize the universality class of the process. They are not parameters subject to variation in entropy maximization.

freedom. This classification scheme is based on *mathematical* analysis of the asymptotic scaling of the entropic functionals that are governed by the first three SK axioms.[2]

Since the configuration space of most complex systems does not grow exponentially (as for the case of Shannon entropy), but polynomially [7], as a stretched exponential [21], or even super-exponentially [22], the appropriate scaling behavior of the entropic functional is crucial for a proper thermodynamic interpretation. For this end, we use recently developed *scaling expansion* [23], which is a special case of Poincaré asymptotic series [24], whose coefficients are the scaling exponents of the system.

The aim of this paper is to define a generalization of Shannon entropy that matches the appropriate asymptotic scaling of a given system, and use it to derive the associated generalized Fisher-Rao metric of the underlying statistical manifold. For this end, we use the framework of deformed logarithms [26,36]. It has been shown recently [27] that one can naturally obtain two types of information metric within that framework, one, corresponding to the maximum entropy principle with linear constraints, and the other, corresponding to the maximum entropy principle when used with so-called escort constraints, instead of ordinary (linear) constraint terms.

Escort distributions appeared in connection with chaotic systems [28], and were discussed in the context of superstatistics [29,30]. Later it became possible to relate them to linear constraints through a *log-duality* [31]. Interestingly, escort distributions also appear as a canonical coordinate in information geometry [32,33]. In this paper, we use both linear and escort approaches and compare their corresponding metric tensor and its invariants. We focus particularly on the microcanonical ensemble in the thermodynamic limit, since the metric should correspond to the system's asymptotic properties, given by its characteristic structure. Some partial results for the curvature of escort metric were recently obtained in this direction [35]. However, no systematic and analytically expressible results for metric tensor and its scalar curvature have been obtained so far. We show that the curvature of the statistical manifold naturally distinguishes between three types of systems: systems with sub-exponentially growing configuration or sample space (correlated and constrained systems), exponentially growing sample space (equivalent to ordinary multinomial statistics), and super-exponentially growing sample space (e.g. systems that develop emergent structures as they evolve). The vector of scaling exponents plays the role of a set of order parameters, i.e., the distance from the phase transition between sub-exponential and super-exponential phases.

The paper is organized as follows: Section 2 introduces the *scaling expansion* and how to calculate corresponding scaling exponents. We discuss several systems with non-trivial scaling exponents. In the last part of the section, we establish a representation of universality classes for complex systems, by introducing *scaling vectors* and their basic operations. In Section 3, we briefly revisit the results of *information geometry* in the framework of ϕ-deformed logarithms. We focus on information geometry with both linear and escort constraints. The main results of the paper are derived in Section 4, where we define the appropriate generalized logarithm by combining the ϕ-deformation framework and the requirement of asymptotic scaling. The properties of corresponding entropic functionals are discussed. We exemplify the whole approach by the simple, yet very general, class of entropies with one correction term from the scaling expansion and calculate the asymptotic behavior of scalar curvature of the microcanonical ensemble in the thermodynamic limit. The last section draws conclusions. The paper has several appendices that contain several technical details.

---

[2]This does not mean that actual distribution functions that are, say, obtained from the maximum entropy principle must be equi-distributed since the form of the distribution is determined not only by the entropic functional, but also by the constraints.

2 Scaling expansion of the volume of configuration space

The scaling expansion [23] is a method to investigate of the asymptotic scaling behavior of a sample space volume, $W(N)$. Here W is the number of accessible states in a system, and N indicates size of a system.[3] The scaling expansion is a special case of the Poincaré asymptotic series, where the coefficients correspond to the scaling exponents of the system. We introduce the notation for the iterated use of functions, $f^{(n)}(x) = \underbrace{f(\dots f(f(x))\dots)}_{n \text{ times}}$, to define a set of re-scaling

operations, $r_\lambda^{(n)}(x) = \exp^{(n)}(\lambda \log^{(n)}(x))$. This set of re-scaling operations contains the well-known multiplicative re-scaling, $x \mapsto \lambda x$ $(n = 0)$, power rescaling $x \mapsto x^\lambda$ $(n = 1)$, and the additive rescaling $x \mapsto x + \log \lambda$ $(n = -1)$. For each n, $r^{(n)}$ is a representation of the multiplicative group $(\mathbb{R}^+, \times)$, i.e., $r_\lambda^{(n)} \circ r_{\lambda'}^{(n)} = r_{\lambda\lambda'}^{(n)}$. We now investigate how a function, $W(N)$, scales with re-scaling of $N \mapsto r_\lambda^{(n)}(N)$. Note that due to a simple theorem (see Appendix A2 in [23]) the function $z(\lambda)$, defined as $z(\lambda) = \lim_{N\to\infty} \frac{g(r_\lambda^{(n)}(N))}{g(N)}$, must have the form $z(\lambda) = \lambda^c$ for $c \in \mathbb{R} \cup \{\pm\infty\}$ whenever the limit exists. We start with multiplicative scaling $(n = 0)$: The expression $\frac{W(\lambda N)}{W(N)}$ is, according to the theorem, equal to λ^{c_0}. We assume that $W(N)$ is a strictly increasing function, then it follows that $c_0 \geq 0$.[4] It can happen that $c_0 = +\infty$. In that case, the expression grows faster than any polynomial. This problem can be resolved by using $\log^{(l)}(W(N))$ instead of $W(N)$, for an appropriate choice of l. The parameter l is chosen such that $c_0^{(l)}$, corresponding to $\frac{\log^{(l)} W(\lambda N)}{\log^{(l)} W(N)} \sim \lambda^{c_0^{(l)}}$, is finite. We call l the *order* of the process. We get that $W(N) \sim \exp^{(l)}(N^{c_0^{(l)}})$, for $N \gg 1$. To get the corrections to the leading order, we use the fact that $\frac{\log^{(l)} W(\lambda N)}{\log^{(l)} W(N)} \frac{N^{c_0}}{(\lambda N)^{c_0}} \sim 1$. When we use the re-scaling for $n = 1$, we get the second scaling exponent: $\frac{\log^{(l)} W(N^\lambda)}{\log^{(l)} W(N)} \frac{N^{c_0^{(l)}}}{(N^\lambda)^{c_0^{(l)}}} \sim \lambda^{c_1^{(l)}}$. Therefore, $W(N) \sim \exp^{(l)}(N^{c_0^{(l)}}(\log N)^{c_1^{(l)}})$. One can continue along the same lines to obtain the asymptotic expansion of $W(N)$, which reads

$$W(N) \sim \exp^{(l)}\left(\prod_{j=0}^{n}(\log^{(j)} N)^{c_j^{(l)}}\right) \quad \text{for } N \to \infty, \tag{1}$$

where $c_j^{(l)}$ are the characteristic scaling exponents. The scaling expansion of $\log^{(l)} W(N)$ can be written

$$\log(\log^{(l)} W(N)) = \sum_{j=0}^{n} c_j^{(l)} \log^{(j+1)} N + \mathcal{O}(\log^{n+1}(N)). \tag{2}$$

It can be shown that the scaling exponents can be calculated from $W(N)$ as

$$c_k^{(l)} = \lim_{N\to\infty} \log^{(k)}(N)\left(\log^{(k-1)}(N)\left(\dots\left(\log(N)\left(N\frac{\mathrm{d}\log^{(l)}(W(N))}{\mathrm{d}N} - c_0^{(l)}\right) - c_1^{(l)}\right)\dots\right) - c_{k-1}^{(l)}\right). \tag{3}$$

---

[3] For example, think of N as the number of particles in a system, or the number of throws in a coin tossing experiment.

[4] Details about processes with reducing sample space can be found e.g., in references [41–44].

As a next step, we apply the scaling expansion to obtain the corresponding *extensive* entropy functionals. It is well-known that for complex systems (with sub- or super-exponential phase space growth) the Shannon-Boltzman-Gibbs entropy is not an extensive quantity. To obtain an extensive expression for such systems, one can introduce an appropriate generalization of the entropy functional [1]. A natural way how to characterize thermodynamic entropy is to define the entropy functional $S(W)$ which is *extensive*. This requirement can be expressed for the microcanonical ensemble as $S(W(N)) \sim N$ for $N \to \infty$. For the purpose of thermodynamics, we do not have to require exact extensivity (with equality sign), but only its weaker asymptotic version. We consider the general trace-form entropy functional

$$S(P) = \sum_{i=1}^{W} g(p_i). \qquad (4)$$

The scaling expansion of the extensive entropy in the microcanonical ensemble can be expressed as

$$S(W) \sim \prod_{j=0}^{n} \log^{(j+l)}(W(N))^{d_j^{(l)}} \qquad \text{for } N \to \infty, \qquad (5)$$

and the scaling expansion of $g(x)$ is

$$g(x) \sim x \prod_{j=0}^{n} \log^{(j+l)} \left(\frac{1}{x} \right)^{d_j^{(l)}}. \qquad (6)$$

The scaling coefficients $d_j^{(l)}$ can be obtained by

$$d_k^{(l)} = \lim_{N \to \infty} \log^{(l+k)}(W) \left(\log^{(l+k-1)}(W) \left(\dots \left(\log^{(l)}(W) \left(\left(N \frac{\mathrm{d} \log^{(l)} W(N)}{\mathrm{d}N} \right)^{-1} \right) - d_0^{(l)} \right) \dots \right) - d_{k-1}^{(l)} \right). \qquad (7)$$

The requirement of extensivity determines the relation between scaling exponents $c_j^{(l)}$ and $d_j^{(l)}$ as

$$d_0^{(l)} = \frac{1}{c_0^{(l)}}$$

$$d_k^{(l)} = -\frac{c_k^{(l)}}{c_0^{(l)}} \quad k = 1, 2, \dots \qquad (8)$$

Examples of systems with different scaling exponents. The first example is a random walk (RW) on the discrete one-dimensional lattice with two possible steps: left or right. The space of all possible paths grows exponentially, $W_{RW}(N) = 2^N \sim \exp(N)$, and we obtain the formula for Boltzmann entropy $S_{RW} = \log W_{RW}$ ($k_B = 1$). Now consider an aging random walk (ARW) [20], where the walker takes one step in a random direction, followed by two steps into a random direction, followed by three steps, etc. In this case, the sample space grows sub-exponentially, $W_{ARW} \sim 2^{\sqrt{N}/2}$, and $S_{ARW} = (\log W_{ARW})^2$. The next example is the magnetic
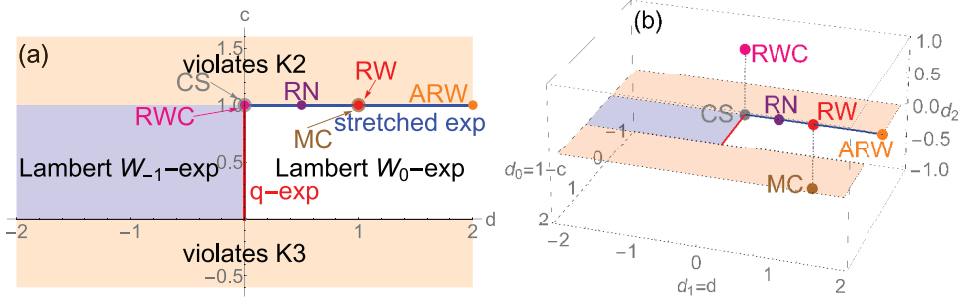
Fig. 1. Parametric space of scaling expansion universality classes with scaling exponents of Random walk (RW), Aging random walk (ARW) Magnetic coin model (MC), Random networks (RN), Random walk cascade (RWC) and processes with compact support distribution (CS). (a) 2D parametric space of scaling expansion universality classes for the first two exponents (as in [19]). We see that some super-exponential systems are not properly represented. (b) Extension to three dimensions by adding the third scaling exponent, d_2. All mentioned examples can be described with the first three scaling exponents.

coin model (MC) [22], where each coin can be in two states: head or tail, however, two coins can also stick together and create a bond state. It can be shown that the corresponding sample space grows super-exponentially, $W_{MC} \sim N^{N/2}e^{2\sqrt{N}}$. One can conclude that the corresponding extensive entropy is asymptotically equivalent to $S_{MC} = \log W_{MC}/\log\log W_{MC}$. Another example of super-exponential processes are random networks (RN), whose sample spaces grow as $W_{RN} = 2^{\binom{N}{2}}$, and thus, $S_{RN} = (\log W_{RN})^{1/2}$. The final example is the double-exponential growth of random walk cascade (RWC), where the walker can take a step to the right, to the left, or split into two independent walkers [23]. For this, we get that $W_{RWC} = 2^{2^N} - 1$, and, $S_{RWC} = \log\log W_{RWC}$. In Figure 1 we show the parameter space of entropies given by three scaling exponents (d_0, d_1, d_2). The above examples are indicated as points. In Figure 1a the plane for the first two scaling exponents is shown, as presented in [19]. We see that if one uses only the first two exponents, some super-exponential processes are not properly represented. By adding a third scaling exponent this problem is solved Figure 1b. So far, we have not yet found simple examples that need more than three scaling exponents.

2.1 Universality classes for scaling expansions

Scaling expansions define universality classes of statistical complex systems according to set of the scaling exponents of their sample space [23]. The representation of the sample space volume, $W(N)$, by its scaling expansion can be used to uniquely describe the statistical properties in the thermodynamic limit.

Consider a function $c(x)$ represented by its scaling expansion

$$c(x) \sim \exp^{(l)}\left[\prod_{j=0}^{n}\left(\log^{(j)}(x)\right)^{c_j^{(l)}}\right].$$ (9)

Its scaling exponents can be collected in the *scaling vector*

$$\mathcal{C} = \{l; c_0^{(l)}, c_1^{(l)}, \ldots, c_n^{(l)}\}.$$ (10)

In principle, the scaling vector can be infinite, however, typically, after several terms the corrections are either zero, or do not contribute significantly. The parameter n denotes the number of corrections.

Let $a(x)$ and $b(x)$ be two functions with its respective scaling expansion determined by the two vectors of scaling exponents

$$\mathcal{A} = \{l_a; a_0, a_1, \ldots, a_n\} \tag{11}$$

$$\mathcal{B} = \{l_b; b_0, b_1, \ldots, b_n\}. \tag{12}$$

Without loss of generality, n can be the same for both vectors because one can always append zeros to the shorter vector. We can now define the *equivalence relation*

$$a(x) \sim b(x) \text{ if } \mathcal{A} \equiv \mathcal{B}, \tag{13}$$

as well as natural ordering

$$a(x) \prec b(x) \text{ if } \mathcal{A} < \mathcal{B}, \tag{14}$$

where the symbol $<$ is used in the lexicographic meaning, i.e.,

$$\mathcal{A} < \mathcal{B} \quad \text{if} \quad \begin{cases} l_a < l_b \\ l_a = l_b, a_0 < b_0 \\ l_a = l_b, a_0 = b_0, a_1 < b_1 \\ \ldots \end{cases} \tag{15}$$

For every vector $\mathcal{C}$ we define the corresponding *entropy scaling vector* $\mathcal{D}$, denoted by $\mathcal{D} = \mathcal{C}^{-1}$, that is obtained from equation (8) by requirement of *extensivity*. One can define analogous relations for $\mathcal{D}$ through the relations for corresponding vectors $\mathcal{C}$. Thus, for entropy scaling vectors $\mathcal{E}$ and $\mathcal{F}$, we can say that

$$\mathcal{E} < \mathcal{F} \quad \text{if} \quad \begin{cases} l_e < l_f \\ l_e = l_f, e_0 < f_0 \\ l_e = l_f, e_0 = f_0, e_1 > f_1 \\ \ldots \end{cases} \tag{16}$$

Note that for sub-leading scaling exponents the inequality is reversed, which is the result of equation (8). Additionally, one can define basic algebraic operations on the scaling vectors, such as generalized addition or derivative operator. More details can be found in Appendix A. Let us make an important note. As discussed in [23], the SK axioms set requirements on the admissible set of scaling exponents. From SK2 we get that $d_l \equiv d_0^{(l)} > 0$ and from SK3 that $d_0 < 1$. Note that the vector $\mathcal{D}$ can be also represented as

$$\mathcal{D} = \{l; d_0^{(l)}, d_1^{(l)}, \ldots, d_n^{(l)}\} = \{\underbrace{0, \ldots, 0}_{l \text{ times}}, d_l, d_{l+1}, \ldots, d_{l+n}\}. \tag{17}$$

This means that one can use the representation without specifying l with an appropriate number of zeros at the beginning. This is useful for example for the plots in the parametric space, where it is possible to plot processes of different order l (as e.g., in Fig. 1). However, one has to keep in mind that this representation can be misleading in the sense that the limit $d_l \to 0$ does not have a clear meaning, since it

changes the order of the process. This can be nicely seen in the example of Tsallis entropy [6], where

$$\lim_{q \to 1^-} \frac{\sum_{i=1}^{W} p_i^q - 1}{1 - q} = -\sum_{i=1}^{W} p_i \log p_i \,, \tag{18}$$

which can be formulated in terms of entropy scaling vectors for as

$$\lim_{q \to 1^-} \mathcal{D} = \lim_{q \to 1^-} (1 - q, 0) = (0, 1). \tag{19}$$

Interestingly, the limit from above, $q \to 1^+$, is even more pathological. In this case the scaling vector corresponding to $S_q(P)$ for $q > 1$ is $(0,0)$, because $S_q(N) \sim N^{1-q} + 1 \sim N^0$. These pathologies have their origin in the non-commutativity of limits, $\lim_{N \to \infty} \lim_{d_l \to 0} \neq \lim_{d_l \to 0} \lim_{N \to \infty}$. The limit $d_l \to 0$ depends on the particular representation of the extensive entropy.

3 Information geometry of ϕ-deformations

Information geometry plays a central role in the theory of information as well as in statistical inference. It allows one to study the structure of the statistical manifold by means of differential geometry. We derive the information-geometric properties of the scaling expansion in the framework of ϕ-deformed logarithms introduced in [26,36]. The ϕ-deformation is a generalization of logarithmic functions. It can be subsequently used to establish a connection with information theory, where the logarithm plays the role of a natural information measure (Hartley information). The ϕ-deformed logarithm is defined by a positive, strictly increasing function $\phi(x)$, on $(0, +\infty)$ as

$$\log_\phi(x) = \int_1^x \frac{\mathrm{d}y}{\phi(y)}. \tag{20}$$

Hence, $\log_\phi$ is an increasing concave function with $\log_\phi(1) = 0$. For $\phi(x) = x$, we obtain the ordinary logarithm. Naturally,

$$\frac{\mathrm{d}\log_\phi(x)}{\mathrm{d}x} = \frac{1}{\phi(x)}. \tag{21}$$

The inverse function of $\log_\phi$, the so-called ϕ-exponential, is an increasing and convex function. This enables one to define the parametric ϕ-exponential family of probability distributions as

$$p(x; \theta) = \exp_\phi \left(\Psi(\theta) + \sum_i x_i \theta_i \right) , \tag{22}$$

where the function $\Psi(\theta)$ is called the *Massieu function* and normalizes the distribution. As discussed in [27], there are two natural ways how to make a connection with the theory of information through the maximum entropy principle. The first is based on the maximization of the entropy functional under the linear (thermodynamic) constraints, the latter is based on a maximization under so-called escort (or geometric)

constraints. Both approaches lead to the ϕ-exponential family. The former approach defines the ϕ-deformed entropy as [26]

$$S_\phi^N(p) = \sum_{i=1}^{W} \int_0^{p_i} \mathrm{d}x \, \log_\phi(x) \tag{23}$$

which is maximized by the ϕ-exponential family for linear constraints, i.e., constraints of the type

$$\sum_{i=1}^{W} p_i E_i = \langle E \rangle. \tag{24}$$

In information geometry, escort distributions play a special role of dual coordinates on statistical manifolds [34]. They can be defined by ϕ-deformations as

$$P_i^\phi = \frac{\phi(p_i)}{\sum_k \phi(p_k)} = \frac{\phi(p_i)}{h_\phi(P)} \ . \tag{25}$$

It can be shown that the entropy maximized by the ϕ-exponential family for escort constraints, i.e., for constraints of the type

$$\sum_{i=1}^{W} P_i^\phi E_i = \langle E \rangle_\phi \ , \tag{26}$$

can be expressed as

$$S_\phi^A(p) = \sum_{i=1}^{W} P_i^\phi \log_\phi(p_i) = \frac{\sum_{i=1}^{W} \phi(p_i) \log_\phi(p_i)}{\sum_{j=1}^{W} \phi(p_j)} \ . \tag{27}$$

For both approaches it is possible to establish a link to information geometry, i.e., to derive a generalization of a Fisher information metric, which can be done through a divergence (or relative entropy) of Bregmann type, which is defined as

$$D_f(p||q) = f(p) - f(q) - \langle \nabla f(q), p - q \rangle \ , \tag{28}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product. Alternatively, one can use the divergence of Csiszár type, but its information geometry is trivial, because it is conformal to ordinary Fisher information geometry, see, e.g., references [27,36].

Let us consider a parametric family of distributions $p(\theta)$. The Fisher information metric of this family at point θ_0 can be calculated as

$$g_{ij}^f(\theta_0) = \frac{\partial^2 D_f(p(\theta_0)||p(\theta))}{\partial \theta_i \partial \theta_j}|_{\theta=\theta_0} \ . \tag{29}$$

Let us consider a discrete probability distribution $\{p_i\}_{i=0}^n$. The normalization is given by $\sum_{i=0}^n p_i = 1$, so we consider $p_i, \ldots, p_n$ as independent variables, while p_0 is determined from $p_0 = 1 - \sum_{i=1}^n p_i$. We parameterize this probability simplex by a ϕ-deformed exponential family.[5] For the entropy S_ϕ^N, we have $f_\phi^N(p) =$

---
[5]Note that this parametric family typically constitutes a smooth manifold [34].

$\sum_i \int_0^{p_i} \log_\phi(x)\, dx$ while for $S_\phi^A(p)$ we end with $f(p) = \sum_i P_i^\theta \log_\phi(p_i)$. After a straightforward calculation, we obtain that [27]

$$g_{\phi,ij}^N(P) = \log_\phi'(p_i)\delta_{ij} + \log_\phi'(p_0) = \frac{1}{\phi(p_i)}\delta_{ij} + \frac{1}{\phi(p_0)} , \tag{30}$$

and

$$g_{\phi,ij}^A(P) = -\frac{1}{h_\phi(p)}\left(\frac{\log_\phi''(p_i)}{\log_\phi'(p_i)}\delta_{ij} + \frac{\log_\phi''(p_0)}{\log_\phi'(p_0)}\right) = \frac{1}{h_\phi(p)}\left(\frac{\phi'(p_j)}{\phi(p_j)}\delta_{ij} + \frac{\phi'(p_0)}{\phi(p_0)}\right) , \tag{31}$$

respectively. As a result, for a given ϕ-deformation there are two types of metric on the information manifold. Note that it is natural to consider a one-parametric class of affine connections for which we obtain the so-called *dually-flat structure* for which the corresponding Christoffel coefficients vanish [34]. This structure is useful in information geometry, however, we stick to the well-known Levi-Civita connection (which can be obtained as a special case of a dually-flat connection, since the Levi-Civita connection is the only self-dual connection [4]), because the metric is non-vanishing. Thus, the corresponding invariants, such as scalar curvature, are non-trivial and reveal some information about the statistical manifold.

Let us now focus on the scalar curvature of corresponding to the metric tensor, $R_\phi = g_\phi^{ik} g_\phi^{lj} R_{\phi,ilkj}$, in the thermodynamic limit $N \to \infty$. We focus on the micro-canonical ensemble, i.e., we consider $p_i = 1/W$. We assume no prior information about the system or its dynamics, so all states are equally probable.

It is possible to show in a technical but straightforward calculation that the scalar curvature can be expressed as (see also [39,40])

$$R_\phi(W) = \frac{W(W-1)}{(2r_\phi(W+1))^2} , \tag{32}$$

which corresponds to the scalar curvature of a W-dimensional ball of radius $2r_\phi$. The function r_ϕ depends only on the form of the ϕ-deformation. We call the function r_ϕ *characteristic length*. For the case of the Amari metric, it can be expressed as

$$(r_\phi^A(W))^2 = -\frac{\log_\phi'\left(\frac{1}{W}\right)^2 \log_\phi''\left(\frac{1}{W}\right)^3}{\left(\log_\phi'''\left(\frac{1}{W}\right)\log_\phi'\left(\frac{1}{W}\right) - 3\log_\phi''\left(\frac{1}{W}\right)^2\right)^2} , \tag{33}$$

while for the metric of Naudts type we obtain

$$(r_\phi^N(W))^2 = \frac{W(\log_\phi'\left(\frac{1}{W}\right))^3}{(\log_\phi''\left(\frac{1}{W}\right))^2}. \tag{34}$$

4 Information geometry of scaling expansions

Let us now consider an arbitrary ϕ-deformed logarithm. We show how to introduce a generalization of the logarithm with a given asymptotic scaling. In contrast to ϕ-deformations, we do not start with the definition of ϕ, but focus on the definition of the logarithm. We denote the desired logarithmic function as $\Lambda_\mathcal{D}$. Let us state the requirements that $\Lambda_\mathcal{D}$ should fulfil:

1. *Domain*: $\Lambda_{\mathcal{D}} : \mathbb{R}^+ \to \mathbb{R}$,

2. *Monotonicity*: $\Lambda'_{\mathcal{D}}(x) > 0$,

3. *Concavity*: $\Lambda''_{\mathcal{D}}(x) < 0$,

4. *Normalization*: $\Lambda'_{\mathcal{D}}(1) = 1$,

5. *Self-duality*: $\Lambda_{\mathcal{D}}(1/x) = -\Lambda_{\mathcal{D}}(x)$,

6. *Scaling expansion*: $\Lambda_{\mathcal{D}}(x) \sim \prod_{j=0}^{k} \left[\log^{(j+l)}(x) \right]^{d_j^{(l)}}$ for $x \to \infty$.

The requirements follow the properties of the ordinary logarithm. Particularly convenient is the self-duality requirement, from which we can directly calculate the asymptotic expansion around 0^+. A direct consequence of self-duality is that $\Lambda_{\mathcal{D}}(1) = 0$. Next, we want to find a representation that is simple, analytically expressible, and universal for any set of scaling exponents. Due to the self-duality requirement, we can focus only on the interval $(1, +\infty)$, while on the interval $(0, 1)$ the logarithm is defined by the self-duality. To find an appropriate representation, we start from the scaling expansion itself. Unfortunately, the scaling expansion, $\prod_{j=0}^{k} \left[\log^{(j+l)}(x) \right]^{d_j^{(l)}}$, is not generally defined on the whole interval $(1, \infty)$, since the domain of $\log^{(l)}(x)$ is $(\exp^{(l-2)}(1), \infty)$. We can overcome this issue by adjusting the nested logarithm by replacing $\log \mapsto 1 + \log$. Further, to be able to fulfil the normalization condition, we add a multiplicative constant to the first nesting, so that for each order the corresponding term can be expressed as $(1 + r_j \log([1 + \log]^{(j-1)}(x))$. Thus, the generalized logarithm can be expressed as

$$\Lambda_{\mathcal{D}}(x) = R \left[\prod_{j=0}^{n} \left(1 + r_j \log[1 + \log]^{(j+l-1)}(x) \right)^{d_j^{(l)}} - 1 \right]. \tag{35}$$

The logarithm automatically fulfils the condition $\Lambda_{\mathcal{D}}(1) = 0$. The parameters r_n define the set of scale parameters that influence the behavior at finite values, while the asymptotic properties are preserved. Because

$$\Lambda'_{\mathcal{D}}(1) = R \sum_{j=0}^{n} r_j d_j^{(l)} \tag{36}$$

we can obtain normalization of the derivative in several ways. For this we define the "calibration"

$$r_0 = \rho \frac{1 - r \sum_{j=1}^{n} d_j^{(l)}}{r d_0^{(l)}} \tag{37}$$

$$r_k = \rho \tag{38}$$

$$R = r/\rho \,, \tag{39}$$

where r and ρ are free parameters. The parameter ρ can be determined by additional requirements. The first option is to require that $\Lambda_{\mathcal{D}}$ is smooth enough, at least it

has continuous second derivative. From the second derivative of the self-duality condition together with the normalization condition, we get $\Lambda''_{\mathcal{D}}(1) = -1$. Following a straightforward calculation, we find

$$\Lambda''_{\mathcal{D}}(1) = R \left(2 \sum_{i<j} r_i r_j d_i^{(l)} d_j^{(l)} + \sum_{j=0}^{n} \left(r_j^2 d_j^{(l)} \cdot \left(d_j^{(l)} - 1 \right) - (j+l) r_j d_j^{(l)} \right) \right). \qquad (40)$$

Using equation (37) in equation (40), we get an expression for ρ_C, i.e., the scale parameter in the smooth calibration

$$\rho_C = \frac{r \sum_j d_j^{(l)} + l - 1}{\frac{d_0^{(l)}-1}{d_0^{(l)} r} \left(1 - r \sum_j d_j^{(l)} \right)^2 + (2-r) \sum_j d_j^{(l)} - 2r \sum_{j \leq i} d_i^{(l)} d_j^{(l)} + r \sum_j \left(d_j^{(l)} \right)^2}. \qquad (41)$$

The free parameter r can be used to ensure that ρ is positive. Alternatively, we can simply consider $r_0 = 1$, which is useful for several applications. In this case we get that ρ_L scale parameter in the leading-order calibration is simply

$$\rho_L = \frac{r d_0^{(l)}}{1 - r \sum_{j=1}^{n} d_j^{(l)}}. \qquad (42)$$

Note that after a proper normalization, this calibration corresponds to the calibration used in [19,20]. Unless a continuous second derivative is explicitly required, it is more convenient to work with this simpler calibration.

We now turn our attention to the information geometry of $\Lambda_{\mathcal{D}}$-deformations and introduce a notation for the nested logarithm

$$\mu_k(x) = [1 + \log]^{(k)}(x). \qquad (43)$$

We sketch the results on for the scaling expansion with one correction. All technical details can be found in Appendix B. In Appendix C we show the calculation for arbitrary scaling vectors and calibrations, which is technically more difficult, but leads to the same results. We now denote the scaling vector as $\mathcal{D} = (l; c, d)$. Note that this entropy has been studied for $l = 0$ in [19]. This inspires us to define the generalized logarithm as

$$\log_{(l;c,d)}(x) = r \left(\mu_l(x)^c \left(1 + \frac{1 - cr}{dr} \log \mu_l(x) \right)^d - 1 \right) = \log_{(c,d)} (\mu_l(x)). \qquad (44)$$

This definition corresponds to the choice of ρ in equation (42).[6] The logarithms are depicted in Figure 2a for various scaling exponents. The inverse function, the deformed exponential, can be obtained in terms of the Lambert W-function[7]

$$\exp_{(l;c,d)}(x) = \nu_l \left(\exp \left(-\frac{d}{c} \left[W \left(B(1 - x/r)^{1/d} \right) - W(B) \right] \right) \right), \qquad (45)$$

---

[6]Note that the original (c,d)-logarithm (as appearing in the rightmost part of Eq. (44)) was introduced in [19] for $l = 0$ and $c \mapsto 1 - c$. Nevertheless, that choice of parametrization is not so convenient for $l > 0$.

[7]The Lambert-W function is defined as the solution of the equation $W(z)e^W(z) = z$.
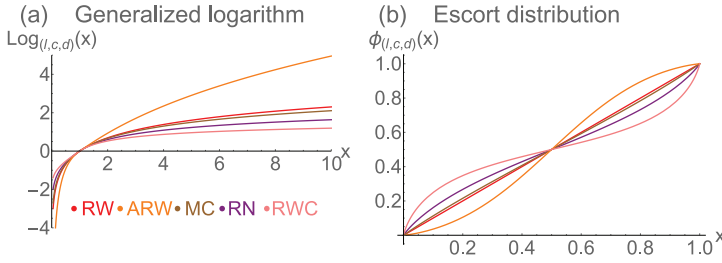
Fig. 2. (a) Generalized logarithms corresponding to scaling exponents of the aforementioned models. (b) Escort distributions corresponding to the generalized logarithms. The scaling exponents $(l; c, d)$ for the models are: Random walk (RW): $(1; 1, 0)$, Ageing random walk (ARW): $(1; 2, 0)$, Magnetic coin model (MC): $(1; 1, -1)$, Random network (RN): $(1; 1/2, 0)$, Random walk cascade (RWC): $(2; 1, 0)$.

where $B = \frac{cr}{1-cr} \exp\left(\frac{cr}{1-cr}\right)$ and ν_l is the inverse function of μ_l, i.e.,

$$\nu_l(x) = \underbrace{\exp(\exp(\ldots(\exp(x - 1) - 1)\ldots))}_{l \text{ times}} .$$ (46)

Note that depending on the values of c and d this deformed exponential contains the exponential, power laws, and stretched exponentials, respectively [1]. It is easy to see that the corresponding scaling vector of the exponential is $\mathcal{C} = (l; 1/c, -d/c)$. The function $\phi_{(l;c,d)}(x)$ can be expressed as

$$\phi_{(l;c,d)}(x) = \phi_{(0;c,d)}(\mu_l(x)) \cdot \mu_l(x)^{-1}$$
$$= \frac{\mu_l(x)}{\log_{(c,d)-r}} \frac{dr + (1 - cr)\log\mu_l(x)}{d + c(1 - cr)\log\mu_l(x)} \prod_{j=0}^{l-1} \mu_k(x) .$$ (47)

The escort distribution, $\rho_{(l;c,d)}(p) = \phi_{(l;c,d)}(p)/(\phi_{(l;c,d)}(p) + \phi_{(l;c,d)}(1 - p))$, corresponding to the two-event distribution $(p, 1 - p)$ is depicted in Figure 2b for various scaling exponents. Interestingly, for $\mathcal{D} < (1; 1)$, i.e., for entropies corresponding to sub-exponential sample space growth, the distribution shows high probabilities (generally $p > 1/N$), while for $\mathcal{D} > (1; 1)$, i.e., for super-exponential growth, the distribution shows low probabilities ($p < 1/N$). Let us finally show the asymptotic behavior of the curvature that corresponds to the deformed logarithm. It can be easily calculated if one keeps only dominant contributions from each term in the equations (33) and (34). In this case we have

$$\frac{d^n \log_{(l;c)}(x)}{dx^n} \approx (\mu_l(x))^{c-1}\mu_l'(x)x^{1-n} \quad \text{for } x \to \infty,$$ (48)

and therefore

$$r_{(l;c)}(W) \approx 1/W \mu_l^{c-1}(W)\mu_l'(W) \quad \text{for } W \to \infty,$$ (49)

for both curvatures, calculated from both types of metric, as shown in Appendix B. From this we deduce that

$$\lim_{W \to \infty} r(W) = \begin{cases} +\infty, & l = 0 \text{ or } l = 1, c > 1 \\ 0, & l \geq 2 \text{ or } l = 1, c < 1 \end{cases} .$$ (50)
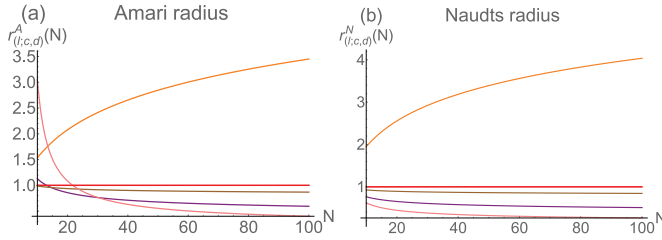
Fig. 3. Characteristic length corresponding to the curvature of the statistical manifold for equiprobable distribution corresponding to different scaling exponents. (a) Corresponds to length of Amari type, (b) corresponds to length of Naudts type, respectively.

For the case $l = 1$ and $c = 1$, we can make a similar approximation

$$\frac{\mathrm{d}^n \log_{(1;1,d)}(x)}{\mathrm{d}x^n} \approx x^{-n} \log(1 + \log(x))^d \,, \tag{51}$$

to get

$$r_{(1;1,d)} \approx \log(1 + \log(W))^d \quad N \to \infty. \tag{52}$$

Similar results can be obtained for higher-order corrections. The behavior of r for different scaling vectors is depicted in Figure 3. We see that the asymptotic behavior is similar for both types of curvature, the only difference is for smaller N.

Similarly, we obtain the same behavior also for higher-order corrections (see Appendix C). In conclusion, we find three distinct regimes for the statistical manifold with respect to the scaling vector

  – $\mathcal{D} < (1;1) \Leftrightarrow r_{\mathcal{D}}(W) \to \infty$ for $W \to \infty$,

  – $\mathcal{D} = (1;1) \Leftrightarrow r_{\mathcal{D}}(W) = 1$ for $W > 0$,

  – $\mathcal{D} > (1;1) \Leftrightarrow r_{\mathcal{D}}(W) \to 0$ for $W \to \infty$.

As a result, the curvature exhibits a phase transition — the statistical manifold in thermodynamic limit is flattening for sub-exponential processes, has constant section curvature for exponential processes, and is curving for super-exponential processes. While processes with exponentially growing sample space have (practically) independent sub-systems, sub-exponential processes impose some restrictions and constaints on the sample space. Super-exponential processes are characterized by emergent structures of its sample space. The scaling vector plays the natural role of the set of order parameters. Let us finally note that the limit $W \to \infty$ is performed for $r_{\mathcal{D}}(W)$. The "limit space" obtained in the limit of the statistical manifolds, for $W \to \infty$, might not be a smooth manifold and the curvature might not correspond to the limit $\lim_{W \to \infty} R_{\mathcal{D}}(W)$.

5 Conclusions and perspectives

In this paper, we have defined a class of deformed logarithms with a given scaling expansion in the framework of ϕ-deformed logarithms. The corresponding entropy can be used to define the statistical manifold with generalized Fisher-Rao metric. We have shown that for the microcanonical ensemble in the thermodynamic limit,

the scalar curvature exhibits a phase transition where the critical point is represented by the class of phenomena that are characterized by exponentially growing phase spaces. These include weakly interacting systems that are correctly described by Shannon entropy. The scaling vector of a given system naturally defines a set of order parameters. A possible explanation for this phenomenon is that the number of independent degrees of freedom grows slower than the size of the system for sub-exponential processes and faster for the super-exponential processes. This classification, however, does not appear for the case of the Fisher metric of Csiszár type, since the characteristic length is constant for every ϕ-deformation.

Contrary to common approach in information geometry, where the statistical manifold corresponds to one functional family of distributions (e.g., exponential family of distributions), this paper presents a parametric way how to switch between different functional families of distributions (e.g., from power-laws to stretched exponentials). This opens a novel connection between parametric and non-parametric information geometry and enables to classify different types of statistical manifolds related to various classes of deformed exponential families.

It will be natural to extend these results to generalizations of Bregmann divergence enabling gauge invariance [45]. Moreover, we will focus on application of the results to the canonical ensemble and use the well-known results using Fisher information metric on the thermodynamic manifold [46,47] for the case of complex systems, where we need to use the generalized form of the Boltzmann factor [48]. Moreover, it should also be possible to go beyond equilibrium statistical mechanics and extend the generalized Fisher metric to non-equilibrium scenarios [49].

Appendix A: Basic algebra of scaling vectors

Let us discuss some definitions of ordinary operations on the space of scaling exponents. First, let us introduce a truncated vector of the scaling vector defined in equation (10) as

$$\mathcal{C}_k = \{l; c_0^{(l)}, c_1^{(l)}, \ldots, c_k^{(l)}\} \tag{A.1}$$

where $k \leq n$. Then, we can introduce

– Truncated equivalence relation: $a(x) \sim^{(k)} b(x)$ if $\mathcal{A}_k \equiv \mathcal{B}_k$

– Truncated inequality relation: $a(x) \prec^{(k)} b(x)$ if $\mathcal{A}_k < \mathcal{B}_k$.

Let us also add one set of inequality relations, and particularly for the case, when even the order l is not equal. For this we define

– Strong inequality relation: $a(x) \ll b(x)$ if $l_a < l_b$.

Let us investigate representations of basic operations on the space of scaling exponents. Before that let us define the rescaling of the general operator $\mathcal{O}$: $\mathbb{R}^m \mapsto \mathbb{R}$ as

$$\mathcal{O}^{(l)}(x_1, x_2, \ldots, x_m) = \exp^{(l)}\left[\mathcal{O}(\log^{(l)} x_1, \log^{(l)} x_2, \ldots, \log^{(l)} x_m)\right]. \tag{A.2}$$

Let us now denote the generalized addition as $a(x) \oplus^{(l)} b(x)$ and multiplication as $a(x) \otimes^{(l)} b(x)$. It is easy to show that

$$a(x) \otimes^{(l)} b(x) = a(x) \oplus^{(l+1)} b(x). \tag{A.3}$$

Let us now consider, without loss of generality, that $a(x) \prec b(x)$. The scaling vector $\mathcal{C}$ of $c(x) = a(x) \otimes^{(l)} b(x)$ can be expressed as follows:

$$\mathcal{C} = \begin{cases} \mathcal{A} + \mathcal{B} = (l, a_0 + b_0, a_1 + b_1, \dots), & \text{for } l_a = l_b = l; \\ \mathcal{B}, & \text{for } l < l_a \leq l_b \text{ or } l = l_a < l_b; \\ \text{undefined}, & \text{for } l > l_a. \end{cases} \tag{A.4}$$

The scaling vector $\mathcal{C}$ of the generalized composition $c(x) = \exp^{(l)} b(\log^{(l)} a(x))$ can be expressed as

$$\mathcal{C} = \begin{cases} b_0^{(l_b)} \mathcal{A} = (l_a + l_b; a_0 b_0, a_1 b_0, a_2 b_0, \dots, a_n b_0), & \text{for } l_a = l; \\ 1^{(l_b)} \mathcal{A} = (l_a + l_b; a_0, a_1, a_2, \dots, a_n), & \text{for } l < l_a; \\ \text{undefined}, & \text{for } l > l_a. \end{cases} \tag{A.5}$$

Finally, let us focus on the derivative of the scaling expansion. Let us denote the rescaled derivative operator as

$$^{(l)} D_x[f] = \exp^{(l)} \left(\frac{\mathrm{d}(\log^{(l)} f(x))}{\mathrm{d}x} \right). \tag{A.6}$$

The scaling vector corresponding to the rescaled derivative is

$$^{(l)} \mathcal{A}' = \begin{cases} (l_a; a_0 - 1, a_1, a_2, \dots, a_n), & \text{for } l_a = l; \\ \mathcal{A}, & \text{for } l_a > l; \\ (l; \underbrace{-1, \dots, -1}_{l - l_a}, 0, \dots), & \text{for } l_a < l. \end{cases} \tag{A.7}$$

Appendix B: Asymptotic curvature of $(l; c, d)$-logarithm

In this Appendix, we calculate asymptotic properties of the $(l; c, d)$ logarithm. Let us first express the derivatives of $(l; c, d)$ logarithm in terms of (c, d) logarithm and μ_l:

$$\log'_{(l;c,d)}(x) = \log'_{(c,d)}(\mu_l(x))\mu'_l(x), \tag{B.1}$$

$$\log''_{(l;c,d)}(x) = \log''_{(c,d)}(x)(\mu'_l(x))^2 + \log'_{(c,d)}(\mu_l(x))\mu''_l(x), \tag{B.2}$$

$$\log'''_{(l;c,d)}(x) = \log'''_{(c,d)}(\mu'_l(x))^3 + 3\log''_{(c,d)}(\mu_l(x))\mu'_l(x)\mu''_l(x)$$
$$+ \log'_{(c,d)}(\mu_l(x))\mu'''_l(x). \tag{B.3}$$

The derivatives of the nested logarithm $\mu_l(x) = [1 + \log]^{(l)}(x)$ can be expressed as:

$$\mu'_l(x) = \frac{1}{\prod_{k=0}^{l-1} \mu_k(x)}, \tag{B.4}$$

$$\mu_l''(x) = -\mu_l'(x) \sum_{k=0}^{l-1} \mu_k'(x) = -\frac{1}{\prod_{k=0}^{l-1} \mu_k(x)} \left(\sum_{k=0}^{l-1} \frac{1}{\prod_{j=0}^{k} \mu_j(x)} \right), \tag{B.5}$$

$$\mu_l'''(x) = -\mu_l''(x) \sum_{k=0}^{l-1} \mu_k'(x) - \mu_l'(x) \sum_{k=0}^{l-1} \mu_k''(x)$$

$$= \mu_l'(x) \sum_{j=0}^{l-1} \mu_j'(x) \sum_{k=0}^{l-1} \mu_k'(x) + \mu_l'(x) \sum_{k=0}^{l-1} \mu_k'(x) \sum_{j=0}^{k-1} \mu_j'(x)$$

$$= 2\mu_l'(x) \sum_{k=0}^{l-1} \mu_k'(x) \sum_{j=0}^{k-1} \mu_j'(x) + \mu_l'(x) \sum_{k=0}^{l-1} \mu_k'(x) \sum_{j=k}^{l-1} \mu_j'(x). \tag{B.6}$$

Let us first denote $l_{(c,d)}(x) = \log_{(c,d)}(x) + r$. Then the derivatives of $\log_{(c,d)}$ can be expressed as (see also Ref. [27]):

$$\log_{(c,d)}'(x) = \frac{l_{c,d}(x)}{x(dr + (1-cr)\log x)} \left[d + c(1-cr)\log x \right], \tag{B.7}$$

$$\log_{(c,d)}''(x) = \frac{l_{c,d}(x)}{x^2(dr + (1-cr)\log x)^2} \left[d\left(d - dr - (cr-1)^2\right) \right.$$
$$\left. + d\left(c^2(r-2)r + 2c - 1\right)\log x + (c-1)c(cr-1)^2 \log^2 x \right], \tag{B.8}$$

$$\log_{(c,d)}'''(x) = \frac{l_{c,d}(x)}{x^3(dr + (1-cr)\log x)^3} \left[d\left(3d(r-1)(cr-1)^2 - 2(cr-1)^3 + d^2\left(2r^2 - 3r + 1\right)\right) \right.$$
$$+ d(cr-1)\left(3c^3r^2 - 3c^2r(r+2) + c\left(d\left(-2r^2 + 6r - 3\right) + 6r + 3\right) + d(3 - 4r) - 3\right)\log x$$
$$- d\left(3c^2(r-1) + c(6 - 4r) - 2\right)(cr-1)^2 \log^2 x$$
$$\left. - c\left(c^2 - 3c + 2\right)(cr-1)^3 \log^3 x \right]. \tag{B.9}$$

In the asymptotic limit, only dominant contributions are relevant. Thus, let us consider only dominant scaling c (i.e., take $d = 0$) and we get

$$\log_{(l;c,d)}'(x) = \log_{(c,d)}'(\mu_l(x))\mu_l'(x) \approx (\mu_l(x))^{c-1}\mu_l'(x), \tag{B.10}$$

$$\log_{(l;c,d)}''(x) \approx \log_{(c,d)}'(\mu_l(x))\mu_l''(x) \approx -\frac{(\mu_l(x))^{c-1}\mu_l'(x)}{x} \tag{B.11}$$

$$\log_{(l;c,d)}'''(x) \approx \log_{(c,d)}'(\mu_l(x))\mu_l'''(x) \approx \frac{(\mu_l(x))^{c-1}\mu_l'(x)}{x^2}. \tag{B.12}$$

Thus, we plugged in equations (33), (34) then we get

$$r_{(l;c)}^A \approx \frac{\left(\frac{\mu_l^{c-1}(x)(\mu_l'(x))}{x} \right) \left(\mu_l^{c-1}(x)(\mu_l'(x)) \right)^2}{\left[\left(\mu_l^{c-1}(x)(\mu_l'(x)) \right) \left(\frac{\mu_l^{c-1}(x)(\mu_l'(x))}{x^2} \right) - 3\left(\frac{\mu_l^{c-1}(x)(\mu_l'(x))}{x} \right)^2 \right]^2} \approx x\mu_l^{c-1}(x)\mu_l'(x), \tag{B.13}$$

and

$$r_{(l;c)}^N \approx \frac{\left(\mu_l^{c-1}(x)(\mu_l'(x)) \right)^3}{x \left(\frac{\mu_l^{c-1}(x)(\mu_l'(x))}{x} \right)^2} \approx x\mu_l^{c-1}(x)\mu_l'(x). \tag{B.14}$$

Let us then focus on the situation $l = 1$, $c = 1$. In this case the leading order terms cancel and we have to look at the first correction given by the scaling exponent d. In this case

$$\log'_{(1;1,d)}(x) \approx \frac{\log(1 + \log(x))^d}{x}, \tag{B.15}$$

$$\log''_{(1;1,d)}(x) \approx \frac{\log(1 + \log(x))^d}{x^2}, \tag{B.16}$$

$$\log'''_{(1;1,d)}(x) \approx \frac{\log(1 + \log(x))^d}{x^3}. \tag{B.17}$$

So the curvature of both Amari and Naudts type can be asymptotically expressed as

$$r_{(1;1,d)} \approx \log(1 + \log(x))^d. \tag{B.18}$$

Appendix C: Fisher metric and scalar curvature corresponding to general logarithm

Let us now show the full calculation of the scalar curvature corresponding to the $\Lambda_{\mathcal{D}}$-logarithm with arbitrary scaling vector $\mathcal{D}$ and constants r_j. Let us first recall the product rule for higher derivatives. The first three derivatives of a function $\Lambda_{\mathcal{D}}(x) = R\left(\prod_{j=0}^{n} \lambda_j(x) - 1\right)$:

$$\Lambda'_{\mathcal{D}}(x) = R\left(\prod_{j=0}^{n} \lambda_j(x)\right)\left[\sum_{j=0}^{n} \frac{\lambda'_j(x)}{\lambda_j(x)}\right], \tag{C.1}$$

$$\Lambda''_{\mathcal{D}}(x) = R\left(\prod_{j=0}^{n} \lambda_j(x)\right)\left[2\sum_{i<j} \frac{\lambda'_i(x)\lambda'_j(x)}{\lambda_i(x)\lambda_j(x)} + \sum_{j=0}^{n} \frac{\lambda''_j(x)}{\lambda_j(x)}\right], \tag{C.2}$$

$$\Lambda'''_{\mathcal{D}}(x) = R\left(\prod_{j=0}^{n} \lambda_j(x)\right)\left[6\sum_{i<j<k} \frac{\lambda'_i(x)\lambda'_j(x)\lambda'_k(x)}{\lambda_i(x)\lambda_j(x)\lambda_k(x)}\right.$$
$$\left. +3\sum_{i<j} \frac{\lambda''_i(x)\lambda'_j(x) + \lambda'_i(x)\lambda''_j(x)}{\lambda_i(x)\lambda_j(x)} + \sum_{i} \frac{\lambda'''_i(x)}{\lambda_i(x)}\right]. \tag{C.3}$$

The derivatives of λ_j can be expressed by defining function $\mathcal{L}$

$$\mathcal{L}_j(x) = \frac{1}{(1 + r_j \log \mu_{j+l-1}(x)) \prod_{k=0}^{j+l-1} \mu_k(x)} = \frac{\mu'_{j+l}(x)}{(1 + r_j \log \mu_{j+l-1}(x))}. \tag{C.4}$$

Then we can express

$$\lambda'_j(x) = \lambda_j(x)\left(r_j d_j \mathcal{L}_j(x)\right), \tag{C.5}$$

$$\lambda''_j(x) = \lambda_j(x)\left(r_j^2 d_j^2 \mathcal{L}_j^2(x) + r_j d_j \mathcal{L}'_j(x)\right), \tag{C.6}$$

$$\lambda'''_j(x) = \lambda_j(x)\left(r_j^3 d_j^3 \mathcal{L}_j^3(x) + 3r_j^2 d_j^2 \mathcal{L}'_j(x)\mathcal{L}_j(x) + r_j d_j \mathcal{L}''_j(x)\right). \tag{C.7}$$

The derivatives of $\mathcal{L}_j(x)$ can be expressed as

$$\mathcal{L}_j'(x) = -\mathcal{L}_j^2(x)\left(r_j + (1 + r_j\log\mu_{j+l-1}(x))\sum_{k=0}^{l+j-1}\prod_{m=k+1}^{j+l-1}\mu_m(x)\right), \quad \text{(C.8)}$$

$$\mathcal{L}_j''(x) = -2\mathcal{L}_j^3(x)\left(r_j + (1 + r_j\log\mu_{j+l-1}(x))\sum_{k=0}^{l+j-1}\prod_{m=k+1}^{j+l-1}\mu_m(x)\right)^2$$
$$-\mathcal{L}_j^2(x)\left(r_j\sum_{k=0}^{l+j-1}\prod_{m=k+1}^{j+l-1}\mu_m(x)\right.$$
$$\left.+(1 + r_j\log\mu_{j+l-1}(x))\sum_{k=0}^{j+l-1}\sum_{m=k+1}^{j+l-1}\frac{\prod_{p=k+1}^{j+l-1}\mu_p(x)}{\prod_{p'=0}^{m}\mu_{p'}(x)}\right). \quad \text{(C.9)}$$

We can finally rewrite the derivatives of Λ_D as

$$\frac{\mathrm{d}}{\mathrm{d}x}(\Lambda_\mathcal{D}(x)) = R\left(\prod_{j=0}^{n}\lambda_j(x)\right)\left[\sum_{j=0}^{n}\mathcal{C}_j^1(x)\mathcal{L}_j(x)\right], \quad \text{(C.10)}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}(\Lambda_\mathcal{D}(x)) = R\left(\prod_{j=0}^{n}\lambda_j(x)\right)\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\mathcal{C}_{ij}^2(x)\mathcal{L}_i(x)\mathcal{L}_j(x)\right], \quad \text{(C.11)}$$

$$\frac{\mathrm{d}^3}{\mathrm{d}x^3}(\Lambda_\mathcal{D}(x)) = R\left(\prod_{j=0}^{n}\lambda_j(x)\right)\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathcal{C}_{ijk}^3(x)\mathcal{L}_i(x)\mathcal{L}_j(x)\mathcal{L}_k(x)\right], \text{(C.12)}$$

where the coefficients $\mathcal{C}$ can be expressed as

$$\mathcal{C}_i^1(x) = r_i d_i, \quad \text{(C.13)}$$
$$\mathcal{C}_{ij}^2(x) = r_i d_i\left[r_j d_j - \delta_{ij}\mathcal{A}_j(x)\right], \quad \text{(C.14)}$$
$$\mathcal{C}_{ijk}^3(x) = r_i d_i\left[r_j d_j r_k d_k - (\delta_{ij}r_j d_j\mathcal{A}_j(x) + \delta_{ik}r_k d_k\mathcal{A}_k(x)\right.$$
$$\left.+ \delta_{jk}r_j d_j\mathcal{A}_j(x)) - \delta_{ijk}\mathcal{B}_i(x)\right], \quad \text{(C.15)}$$

where

$$\mathcal{A}_i(x) = \left((r_i + (1 + r_i\log\mu_{i+l-1}(x)))\sum_{k=0}^{i+l-1}\prod_{m=k+1}^{i+l-1}\mu_m(x)\right), \quad \text{(C.16)}$$

$$\mathcal{B}_i(x) = 2\mathcal{A}_i(x)^2 + \mathcal{L}_i(x) \left(r_i \sum_{k=0}^{l+j-1} \prod_{m=k+1}^{i+l-1} \mu_m(x) \right.$$

$$\left. + (1 + r_i \log \mu_{i+l-1}(x)) \sum_{k=0}^{i+l-1} \sum_{m=k+1}^{i+l-1} \frac{\prod_{p=k+1}^{i+l-1} \mu_p(x)}{\prod_{p'=0}^{m} \mu_{p'}(x)} \right). \qquad \text{(C.17)}$$

Finally, we plug the expressions for $\Lambda_{\mathcal{D}}$ and its derivatives into equations (33), (34), and we end with

$$r_{\mathcal{D}}^A(x) = \frac{\left(\sum_i \mathcal{C}_i^1 \mathcal{L}_i(x)\right)^2 \left(\sum_{kl} \mathcal{C}_{kl}^2(x) \mathcal{L}_k(x) \mathcal{L}_l(x)\right)^3}{\left(\sum_{ijkl} \left(\left[\mathcal{C}_i^1(x)\mathcal{C}_{jkl}^3(x) - 3\mathcal{C}_{ij}^2(x)\mathcal{C}_{kl}^2(x)\right] \mathcal{L}_i(x)\mathcal{L}_j(x)\mathcal{L}_k(x)\mathcal{L}_l(x) \right)\right)^2}, \quad \text{(C.18)}$$

and

$$r_{\mathcal{D}}^N(x) = \frac{\left(\sum_i C_i^1(x)\mathcal{L}_i(x)\right)^3}{x \left(\sum_{ij} C_{ij}^2(x)\mathcal{L}_i(x)\mathcal{L}_j(x)\right)^2}, \qquad \text{(C.19)}$$

for $x = 1/W$, respectively.

References

1. S. Thurner, R. Hanel, P. Klimek, *Introduction to the Theory of Complex Systems* (Oxford University Press, Oxford, UK, 2018)
2. S. Thurner, B. Corominas-Murtra, R. Hanel, Phys. Rev. E **96**, 032124 (2017)
3. N. Ay, J. Jost, H.V. Le, L. Schwachhöfer, *Information Geometry* (Springer, Berlin, Germany, 2017)
4. S.-I. Amari, *Information Geometry and Its Applications* (Springer, Japan, 2016)
5. W. Janke, D.A. Johnston, R. Kenna, Physica A **336**, 181 (2004)
6. C. Tsallis, J. Stat. Phys. **52**, 479 (1988)
7. C. Tsallis, M. Gell-Mann, Y. Sato, Proc. Natl. Acad. Sci. USA **102**, 15377 (2005)
8. G. Kaniadakis, Phys. Rev. E **66**, 056125 (2002)
9. P. Jizba, T. Arimitsu, Ann. Phys. **312**, 17 (2004)
10. C. Tsallis, L.J. Cirto, Eur. Phys. J. C **73**, 2487 (2013)
11. T.S. Biró, V.G. Czinner, H. Iguchi, P. Ván, Phys. Lett. B **782**, 228 (2018)
12. C.E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948)
13. C.E. Shannon, Bell Syst. Tech. J. **27**, 623 (1948)
14. A.I. Khinchin, *Mathematical Foundations of Information Theory* (Dover, New York, 1957)
15. S. Abe, Phys. Lett. A **271**, 74 (2000)
16. V.M. Ilić, M.S. Stanković, Physica A **411**, 138 (2014)
17. P. Tempesta, Phys. Rev. E **84**, 021121 (2011)
18. P. Jizba, J. Korbel, Phys. Rev. Lett. **122**, 120601 (2019)
19. R. Hanel, S. Thurner, Europhys. Lett. **93**, 2006 (2011)
20. R. Hanel, S. Thurner, Europhys. Lett. **96**, 50003 (2011)
21. C. Anteneodo, A.R. Plastino, J. Phys. A **32**, 1089 (1999)

22. H.J. Jensen, R.H. Pazuki, G. Pruessner, P. Tempesta, J. Phys. A **51**, 375002 (2018)
23. J. Korbel, R. Hanel, S. Thurner, New J. Phys. **20**, 093007 (2018)
24. E.T. Copson, *Asymptotic Expansions, Cambridge Tracts in Mathematics* (Cambridge University Press, 1965)
25. J. Naudts, Physica A **316**, 323 (2002)
26. J. Naudts, *Generalised thermostatistics* (Springer Science & Business Media, 2011)
27. J. Korbel, R. Hanel, S. Thurner, Entropy **21**, 112 (2019)
28. C. Beck, F. Schögl, *Thermodynamics of chaotic systems: an introduction* (Cambridge University Press, 1995)
29. C. Beck, E.D.G. Cohen, Physica A **322**, 267 (2003)
30. C. Tsallis, A.M.C. Souza, Phys. Rev. E **67**, 026106 (2003)
31. R. Hanel, S. Thurner, M. Gell-Mann, Proc. Natl. Acad. Sci. USA **109**, 19151 (2012)
32. S. Abe, Phys. Rev. E **68**, 031101 (2003)
33. A. Ohara, H. Matsuzoe, S.-I. Amari, J. Phys.: Conf. Ser. **201**, 012012 (2010)
34. S.-I. Amari, A. Ohara, H. Matsuzoe, Physica A **391**, 4308 (2012)
35. D.P.K. Ghikas, F.D. Oikonomou, Physica A **496**, 384 (2018)
36. J. Naudts, Rev. Math. Phys. **16**, 809 (2004)
37. F. Caruso, C. Tsallis, Phys. Rev. E **78**, 021102 (2008)
38. J.A. Carrasco, F. Finkel, A. González-López, M.A. Rodríguez, P. Tempesta, J. Stat. Mech. Theor. Exp. **2016**, 033114 (2016)
39. J. Zhang, Neural Comput. **16**, 159 (2004)
40. S.-I. Amari, *Differential-Geometrical Methods in Statistics*. Lecture Notes in Statistics (Springer Science & Business Media, 2012), Vol. 28
41. B. Corominas-Murtra, R. Hanel, S. Thurner, Proc. Natl. Acad. Sci. USA **112**, 5348 (2015)
42. B. Corominas-Murtra, R. Hanel, S. Thurner, New J. Phys. **18**, 093010 (2016)
43. B. Corominas-Murtra, R. Hanel, S. Thurner, Sci. Rep. **7**, 11223 (2017)
44. R. Hanel, S. Thurner, Entropy **20**, 838 (2018)
45. J. Naudts, J. Zhang, Inf. Geometry **1**, 79 (2018)
46. G. Ruppeiner, Rev. Mod. Phys. **67**, 605 (1995)
47. G.E. Crooks, Phys. Rev. Lett. **99**, 100602 (2007)
48. R. Hanel, S. Thurner, Physica A **351**, 260 (2005)
49. S. Ito, Phys. Rev. Lett. **121**, 030605 (2018)