Family of additive entropy functions out of thermodynamic limit

Alexander N. Gorban*

Institute of Computational Modeling RAS, 660036 Krasnoyarsk, Russia

Iliya V. Karlin†

ETH Zürich, Department of Materials, Institute of Polymers

ETH-Zentrum, Sonneggstr. 3, ML J 19, CH-8092 Zürich, Switzerland

(Dated: May 24, 2002)

Abstract

Starting with the additivity condition for Lyapunov functions of master equation, we derive a one-parametric family of entropy functions which may be appropriate for a description of certain effects of finiteness of statistical systems, in particular, distribution functions with long tails. This one-parametric family is different from Tsallis entropies, and is essentially a convex combination of the Boltzmann-Gibbs-Shannon entropy and the entropy function introduced by Burg. An example of how longer tails are described within the present approach is worked out for the canonical ensemble. In addition, we discuss a possible origin of a hidden statistical dependence, and give explicit recipes how to construct corresponding generalizations of master equation.

PACS numbers: 05.70.Ln, 05.20.Dd

*Electronic address: gorban@icm.krasn.ru

†Electronic address: ikarlin@ifp.mat.ethz.ch

I. INTRODUCTION

Past several years have witnessed a burst of interest in nonextensive statistical mechanics, the topic which finds increasingly more applications particularly due to the concept of Tsallis entropy [1], [2]. In this approach, one postulates the following one-parametric family of concave functions,

$$S_q = \frac{1 - \sum_i p_i^q}{1 - q}, \tag{1}$$

where $q > 0$. The family of Tsallis entropies (1) replaces the traditional Boltzmann-Gibbs-Shannon entropy, S_1,

$$S_1 = \lim_{q \to 1} S_q = -\sum_i p_i \ln p_i. \tag{2}$$

One of the important achievements associated with Tsallis entropy (1) is the fact that it provides an easy access - through the method of entropy maximization - to a rich set of distribution functions, different from the traditional Gaussian distribution function, thereby giving a handy description of long (non-exponential) tails of probability distributions. For each value of the parameter q, Tsallis entropy is a concave function of the probability. The characteristic feature of Tsallis entropy is that it is nonextensive for $q \neq 1$, that is, if the system is composed of two statistically independent subsystems, Tsallis entropy of this system is not the sum of Tsallis entropies of the subsystems. Since Tsallis entropy is postulated rather than derived, this point remains somewhat open to discussions [3, 4].

The goal of this paper is to present an argument on how long tails can be described in a usual, extensive (more precisely, almost extensive) statistical mechanics, and to give a theoretical derivation of a different, and in a certain sense unique, one-parametric family of entropy functions which can model effects of finiteness.

Our first remark is that real-world systems, to which statistical mechanics is applied, are finite, and, although they do consist of a large number of subsystems, the natural logarithm of this number (and since we address questions related to entropies, one should observe the magnitude of the logarithm, in the first place) is not that big after all, it is not larger than 100 and is often less than 20. Extensivity, in a true sense of this notion, theorems of equivalence of the microcanonical and the canonical ensembles [5], and like, is valid only in the thermodynamic limit where the system can be partitioned into an arbitrary large number of noninteracting and statistically independent subsystems. Namely, it is the number of such independent subsystems ν which do not interact, and which are similar in all their observable properties to the larger system, that plays the role of the parameter the values of which tells how close is the system to the thermodynamic limit.

One is up to invoking that ν is finite (and by doing so, one restores to an argument about an incomplete extensivity) when one needs to cut off the tails of distribution functions with divergent averages. This fact is well known, for example, in the case of the classical Boltzmann equation: The maximum entropy solution for the Boltzmann entropy does not exist (is not normalizable) if the observables are the density, the average momentum, the stress tensor and the heat flux [6], [7], [8], [9], [10]. A regularization by the argument that the magnitude of the microscopic velocity is restricted to the value dictated by finiteness of the *total* energy [7] is an example of the incomplete extensivity argument.

Thus, when the system is not strictly in the thermodynamic limit, details of the interaction should gradually become more and more important, and prospects of a *universal* description using a maximization of an *interaction-independent* entropy functional become less evident. Nevertheless, the very possibility of a sufficiently reliable universal description in the sense just mentioned cannot be ruled out a priori. For that reason, a search for non-classical entropies for a possible description of nonextensive systems seems motivated.

The structure of this paper is as follows: In the next section II, we shall review, for the sake of completeness, the theory of Lyapunov functions of master equation. In section III we derive the family of the (almost) additive entropies from the condition of additivity of the entropy function for statistically independent systems. In section IV we demonstrate with a simple example how long tails are related to the effects of finiteness in the present approach.

Sections III and IV are the central point of our presentation. In section V we discuss a different scenario how the apparent statistical dependence can occur when description of the system is incomplete, and present a natural generalization of master equation for those cases in section VI. Finally, results are briefly discussed in section VII.

II. LYAPUNOV FUNCTIONS OF MASTER EQUATION

We shall start our discussion with a brief summary of the theory of Markov chains. Our presentation essentially follows Ref. [11]. Let us consider a finite set of states $E_1, \dots, E_N$, and let us assume that the system can be found only in those states. The probability of finding the system at time $t \geq 0$ is given by the N-component vector $\boldsymbol{p}$ of probabilities $p_i(t)$, such that $p_i \geq 0$, and $\sum_{i=1}^{N} p_i = 1$. Dynamic equation for $\boldsymbol{p}$ of the form,

$$\dot{p}_i = \sum_{k=1}^{N} q_{ik} p_k, \tag{3}$$

describes the time evolution of a Markov chain if and only if the matrix elements q_{ik} satisfy $q_{ik} \geq 0$ for $i \neq k$, and for every k,

$$\sum_{i=1}^{N} q_{ik} = 0. \tag{4}$$

For each Markov chain, a graph of transition is put into correspondence by drawing an oriented link from the vertex E_i to the vertex E_k if $q_{ki} > 0$. Important class of Markov chains is characterized by directional connectivity. The graph of transitions is called directionally connected if there is a path from each vertex to any other vertex made up of the oriented links. Then the following ergodic theorem is valid: Let transition graph of the Markov chain be directionally connected. Then there exists a positive stationary state $\boldsymbol{p}^{\mathrm{eq}}$, $p_i^{\mathrm{eq}} > 0$, and for any initial condition $\boldsymbol{p}(0)$, solution $\boldsymbol{p}(t)$ to equation (3) tends to $\boldsymbol{p}^*$ at $t \to \infty$.

Let the Markov chain satisfy to the ergodic theorem, and let the stationary state $\boldsymbol{p}^*$ is known. Let $h(x)$ be a convex and twice differentiable function of one variable $x \in [0, \infty]$. Any function h defines a convex Lyapunov function H_h of the Markov chain (3),

$$H_h(\boldsymbol{p}) = \sum_{i=1}^{N} p_i^{\mathrm{eq}} h(p_i/p_i^{\mathrm{eq}}). \tag{5}$$

Time derivative of the function H_h (5) due to dynamics (3) is nonpositive:

$$\dot{H}_h = \sum_{i,j,i\neq j} q_{ij} p_j^{\mathrm{eq}} [h(p_i/p_i^{\mathrm{eq}}) - h(p_j/p_j^{\mathrm{eq}}) + h'(p_i/p_i^{\mathrm{eq}})((p_j/p_j^{\mathrm{eq}}) - (p_i/p_i^{\mathrm{eq}}))] \leq 0, \tag{6}$$

where prime denotes derivative with respect to argument. Equality sign is achieved only in the stationary state.

The stationary state $\boldsymbol{p}^{\rm eq}$ is called the state of detail balance, if it satisfies

$$q_{ik}p_k^{\rm eq} = q_{ki}p_i^{\rm eq}. \tag{7}$$

Markov chain with detail balance is colloquially termed master equation. In this case, the time derivative of the Lyapunov function becomes especially simple,

$$\dot{H}_h = -\frac{1}{2}\sum_{i,j,i\neq j} q_{ij}p_j^{\rm eq}[h'(p_i/p_i^{\rm eq}) - h'(p_j/p_j^{\rm eq})][(p_i/p_i^{\rm eq}) - (p_j/p_j^{\rm eq})] \leq 0. \tag{8}$$

Discussion of the physical significance of the detail balance for the Markov chain (master equation) is a well known textbook material.

Since a convex linear combination of convex function is also a convex function, the obvious construction which enables one to create new Lyapunov functions from given representatives of the family (5) is this: Let $h_1, \dots, h_k$ be convex functions, and $\alpha_1, \dots, \alpha_k$ be nonnegative numbers, $\alpha_i \geq 0$, satisfying $\sum_{m=1}^k \alpha_m = 1$. Then,

$$H_{\alpha_1 h_1 \dots \alpha_k h_k} = \sum_{m=1}^{k}\sum_{i=1}^{N} p_i^{\rm eq}\alpha_m h_m(p_i/p_i^{\rm eq}), \tag{9}$$

is also the Lyapunov function of Markov chain. It should be stressed that the set (9) does not extends the family (5) already specified.

Concluding this summary, we stress that, under physically significant restrictions on the existence of the stationary state, any Markov chain has a large class of Lyapunov functions of the form (5), each constructed from a convex function h of one variable. The additivity requirement makes it possible to drastically restrict the class of physically relevant Lyapunov functions of Markov chains, which we do in the next section.

III. FAMILY OF ADDITIVE LYAPUNOV FUNCTIONS

In order to derive the family of additive Lyapunov functions, let us consider two statistically independent systems described by probability vectors $\boldsymbol{p}$ and $\boldsymbol{q}$, $p_i \geq 0$, $q_j \geq 0$, $\sum_i p_i = 1$, $\sum_j p_j = 1$. respectively. We shall consider first the case of the equipartition at the equilibrium, in order to simplify notation. Thus, we assume that equilibrium states of both the systems are equipartitions with probability vectors $\boldsymbol{p}^{\rm eq}$ and $\boldsymbol{q}^{\rm eq}$, where $p_i^{\rm eq} = 1/P$, $q_i^{\rm eq} = 1/Q$, and where P and Q are the numbers of states in the systems.

Since the systems are independent, the joint system is characterized by the joint probability vector $\boldsymbol{pq}$. The equilibrium of the joint system is again the equipartition, $(\boldsymbol{pq})^{\mathrm{eq}} = \boldsymbol{p}^{\mathrm{eq}}\boldsymbol{q}^{\mathrm{eq}}$, that is the equilibrium is multiplicative with respect to joining the systems if the latter are statistically independent. The condition of additivity for the Lyapunov function (5) of the joint system reads:

$$H_h(\boldsymbol{pq}) = H_h(\boldsymbol{p}) + H_h(\boldsymbol{q}). \tag{10}$$

This functional equation has two special solutions, corresponding to the convex functions, $h_1(x) = x\ln x$, and $h_2(x) = -\ln x$. We denote $H_1 = H_{x\ln x}$ and $H_2 = H_{-\ln x}$, respectively. Whereas the function H_1 corresponds to the classical (additive) Boltzmann-Gibbs-Shannon entropy, we shall demonstrate here the additivity of H_2. Indeed,

$$\begin{aligned} H_2(\boldsymbol{pq}) &= -\sum_{\{ij\}=1}^{PQ} P^{-1}Q^{-1}\ln(PQp_iq_j) \\ &= -\ln(PQ) - \sum_{i=1}^{P}\ln p_i - \sum_{j=1}^{Q}\ln q_j \\ &= -\sum_{i=1}^{P} P^{-1}\ln(Pp_i) - \sum_{j=1}^{Q} Q^{-1}\ln(Qq_j) \\ &= H_2(\boldsymbol{p}) + H_2(\boldsymbol{q}). \end{aligned}$$

Neglecting the irrelevant constant and constant factors, and using Eq. (9), we finally arrive at the one-parametric family of additive convex Lyapunov functions (5) for master equation with N states:

$$H_\alpha = (1-\alpha)\sum_{i=1}^{N} p_i\ln p_i - \alpha\frac{1}{N}\sum_{i=1}^{N}\ln p_i,\ 0\le\alpha\le 1 \tag{11}$$

The one-parametric family of additive Lyapunov functions is the central point of our further discussion. Several remarks are in order:

Remark 1 In the thermodynamic limit, which in the case considered here corresponds formally to $N\to\infty$, for any α, we have $H_\alpha \to (1-\alpha)H_1$. That is, the non-classical contribution due to H_2 becomes significant only if the system is not too close to the thermodynamic limit. In the thermodynamic limit survives only the classical Boltzmann-Gibbs-Shannon contribution.

Remark 2 It is not difficult to prove that the family (11) exhausts all the possible additive Lyapunov functions of the form (5) (up to adding a constant, and a multiplication with a constant factor): Indeed, the classical treatment of the additivity condition requires averaging the vector function $\ln \boldsymbol{p}$ which can be done either using $\boldsymbol{p}$ or $\boldsymbol{p}^{\mathrm{eq}}$. The latter is the distinguished probability distribution which is, same as $\boldsymbol{p}$, multiplicative with respect to joining the statistically independent subsystems. Relevance of master equation, and hence of kinetic rather than of static picture, to our derivation of the one-parametric family (11) is apparent: This enables to consider *two* sets of probabilities, the "current" $\boldsymbol{p}$, and the "final" $\boldsymbol{p}^{\mathrm{eq}}$ (the equipartition here).

Other convex functions which are additive under joining statistically independent systems do exist, for example, the Rényi entropy function [12] but they are not of the form (5) (that is not of the so-called "trace form", cf. Ref. [13]). For this reason, such functions fall out of our discussion since the proofs of the inequalities (6) and (8) are not valid for them.

Remark 3 Function H_2 is not defined (and, consequently, any of the function H_α, $\alpha \neq 0$ is not defined) if one of the probabilities p_i equals to zero. The classical Boltzmann-Gibbs-Shannon solution to the additivity equation is distinguished by the property of continuity at $p_i = 0$. This is a blueprint of the long-tail features (see next section). Work with the family of entropies (11) assumes preserving additivity on the expense of abandoning the continuity of the entropy functions on closed intervals $0 \leq p_i \leq 1$, and its replacement by continuity on semi-open intervals, $0 < p_i \leq 1$.

Remark 4 To the best of our knowledge, the entropy function

$$S_2 = -H_2 = \sum_{i=1}^{N} \ln p_i \tag{12}$$

has been first considered by Burg in the context of applications of information theory to geophysical problems [14], [15]. Recently, the Burg entropy (12) has been used to construct illustrations of the entropic lattice Boltzmann method [16] in Ref. [17]. However, we failed to find a reference to the one-parametric family (11) prior to Ref. [11]. Whereas in Ref. [11] the one-parametric family (11) has been mentioned as just the solution to the additivity condition, its relevance to describing effects of finiteness in statistical systems has not been duly discussed.

Remark 5 If the equilibrium $\boldsymbol{p}^{\mathrm{eq}}$ of the Markov chain differs from the equipartition but is multiplicative under joining statistically independent subsystems, the one-parametric family

(11) generalizes to the following:

$$H_\alpha = (1-\alpha)\sum_{i=1}^{N} p_i \ln\left(\frac{p_i}{p_i^{\mathrm{eq}}}\right) - \alpha \sum_{i=1}^{N} p_i^{\mathrm{eq}} \ln\left(\frac{p_i}{p_i^{\mathrm{eq}}}\right). \tag{13}$$

IV. LONGER TAILS: AN EXAMPLE

In this section we want to explicitly work out an example in order to demonstrate that the entropies of the family (11) indeed describes the long tails for $\alpha \neq 1$. When the discrete system of states is addressed, as it is done here, the meaning of the long tail has to be understood as broadening of the distribution functions.

Without loss of generality, we shall work in this section with the one-parametric set,

$$H_\alpha = (1-\alpha)\sum_{i=1}^{N} p_i \ln p_i - \alpha \sum_{i=1}^{N} \ln p_i, \;\; 0 \leq \alpha \leq 1. \tag{14}$$

We shall consider first the microcanonic ensemble, that is, the minimizer of H_α under the constraint of fixed normalization, $\sum_{i=1}^{N} p_i = 1$. It is straightforward to see that, for any admissible value of the parameter α, the microcanonic state is the equipartition, as expected.

In order to address the canonic ensemble, we introduce energies of the states $E_i \geq 0$, and find the minimum of H_α (14) under the constraints,

$$\sum_{i=1}^{N} p_i = 1, \tag{15}$$

$$\sum_{i=1}^{N} E_i p_i = U. \tag{16}$$

Denoting the solution $\boldsymbol{p}^{(\alpha)}$, we find, for $\alpha \neq 1$,

$$p_i^{(\alpha)} \exp\left\{-\frac{\alpha}{(1-\alpha)p_i^*}\right\} = \exp\{\lambda - \beta E_i\}, \tag{17}$$

where λ and β are Lagrange multipliers, corresponding to the constraints (15). In order to address the effect of $\alpha \neq 0$, we shall restore to a perturbation theory around the Boltzmann-Gibbs-Shannon point $p_i^{(0)}$. After a few algebra, we get for $\alpha \ll 1$, $\alpha > 0$:

$$p_i^{(\alpha)} = p_i^{(0)} + \alpha\left(1 - Np_i^{(0)} + (U - E_i)\frac{V - NU}{C - U^2} p_i^{(0)}\right), \tag{18}$$

where

$$p_i^{(0)} = \frac{1}{Z^{(0)}} e^{-\beta^{(0)} E_i}, \;\; Z^{(0)} = \sum_{j=1}^{N} e^{-\beta^{(0)} E_j}, \tag{19}$$

is the canonical distribution function for the Boltzmann-Gibbs-Shannon entropy (Lagrange multiplier $\beta^{(0)}$ is expressed in terms of the average energy U by the constraint (16); we do not need here the explicit expression $\beta^{(0)}(U)$), and

$$V = \sum_{i=1}^{N} E_i, \tag{20}$$

$$C = \sum_{i=1}^{N} E_i^2 p_i^{(0)}, \tag{21}$$

that is, V is the total energy of the states, and the denominator appearing in Eq. (18), $C - U^2$, is the correlation of the energy levels E_i in the canonical state (19). We further denote,

$$B = \frac{V - NU}{C - U^2}. \tag{22}$$

It can be argued that $B > 0$: The total energy of the states, V, is not less (and in most of the relevant cases, much larger) than the average energy U times the number of states, whereas the correlator $C - U^2$ is always positive.

Function (18) is the first-order perturbation result, and it is not a positive definite quantity. Yet, it is sufficient to our purpose here, since the question we want to address is as follows: What is the sign of the derivatives, $dp_i^{(\alpha)}/d\alpha\big|_{\alpha=0}$? Probabilities of the canonical distribution (19) decay when E_i exceed the average energy U, so, by switching on the Burg component, do we see the "raising" of the populations of this "high-energy tail"? In order to see this, we obtain in Eq. (17),

$$\left.\frac{dp_i^{(\alpha)}}{d\alpha}\right|_{\alpha=0} = A_i p_i^{(0)}, \tag{23}$$

where the factor A_i is,

$$A_i = Z^{(0)} e^{\beta^{(0)} E_i} - N - B(E_i - U). \tag{24}$$

Factor A_i amplifies populations of the states which are less populated in the standard canonical ensemble (19) if E_i satisfies the inequality, $E_i > \epsilon$, where ϵ is the solution to the equation:

$$(1/Z^{(0)}) e^{-\beta^{(0)}\epsilon} (N + B(\epsilon - U)) = 1. \tag{25}$$

In order to make the situation even more transparent, we shall assume that the energies E_i are in a narrow band around the value $E > 0$, that is $E_i = E + \delta_i$, $\sum_{i=1}^N \delta_i = 0$, and $\delta_i \ll E$. All the quantities contributing to the expression (18) can be then evaluated in terms of expansion in δ_i (notice that the second-order perturbation in δ_i must be used in order to compute the correlation $C - U^2$). We obtain, $B = \beta_E^{(0)} + o(\delta_i^2)$, and, up to second order, eq. (25) reads

$$\beta(N/2 - 1)\delta^2 + \beta(1 - N)\delta + \beta^2(N^{-1} - 1/2)\sum_{i=1}^N \delta_i^2 = 0. \tag{26}$$

For large N, this gives that factor (24) is larger than zero, and hence amplifies the populations of the energy levels $E + \delta_i$, if

$$\delta_i \geq 2/\beta_E^{(0)}. \tag{27}$$

That is, raising of the populations of higher energy levels is explicitly demonstrated in this example. We do not discuss corrections for finite N to the estimate (27) which are easily obtained from Eq. (26).

Thus, we have demonstrated with explicit example that taking into account the Burg component in the one-parametric family (11) indeed is able to describe broadening of the canonical distribution function. Appearance of the energy levels correlation in the above formula (18) remarkably resembles recent results of application of Tsallis entropy to fitting experimental data in turbulence (see Ref. [18] and references cited therein).

Generalizations to quasi-equilibrium situation with more constraints is straightforward. Also, if generalizations to a continuous case of states are addressed, the long tail feature of the corresponding distribution functions becomes even more apparent. For example, the counterpart of the Gaussian distribution function $P(x)$ (maximizer of the Boltzmann-Gibbs-Shannon entropy under the constraint fixing the normalization and the variance) has the form, $P(x) \sim (\lambda + \beta x^2)^{-1}$. For it, algebraic decay at infinity precludes existence of the normalization, and a cutoff is required.

To this end, we have argued that for systems out of the strict thermodynamic limit, there exists a universal (that is, not explicitly dependent on details of interactions) one-parametric family of additive entropy functions, which are able to describe, at least in principle, the same long-tail effects as the Tsallis entropy. In the remainder of this paper we shall discuss a different issue of how additivity of the entropy can *apparently* be violated if the description of the system is incomplete.

V. NON-ADDITIVITY AND INCOMPLETE DESCRIPTION

Discussions aimed at justifying a non-additive dependence $S(\boldsymbol{p})$ are sometimes conducted in a rather obscure way: One argues that the entropy is non-additive under joining the statistically independent subsystems because, *in reality*, these subsystems are not independent. Possible physical agents which could lead to such a situation are occasionally mentioned, like long-range forces, for example, which seemingly justify statements like "the concept of *independent* subsystems does not make any sense, since all subsystems are interacting" [18]. However, to the best of our knowledge, a direct demonstration has never been done for any realistic system. Moreover, one should be more cautious in rejecting the "concept of independent subsystems", especially when discrete systems as above are addressed, since this may lead to a confrontation with the traditional axiomatic (Kolmogorov's) probability theory which it is strongly based on this concept [19] (in the worst case, one should abandon the concept of independent trials which is at the very hart of the definition of probabilities).

This all leads to a question: If the probability distribution over pairs of states p_{ij} factors into products of distributions, $p_{ij} = q_i r_j$, but subsystems are dependent, then where this dependence is hidden? What is this "new" notion of independence which does not coincide with the classical, $\mathrm{Prob}(A|B) = \mathrm{Prob}(A)$ *means* A is independent of B? In order to answer this question, one should realize that such a situation of a "hidden" dependence, in fact, has been long known in physics. This is the Pauli exclusion principle. The corresponding Fermi-Dirac entropy has the well known form:

$$S(\boldsymbol{p}) = -\sum_i [p_i \ln p_i + (1-p_i)\ln(1-p_i)]. \tag{28}$$

This expression can be interpreted in the following way: With the electron gas, there is associated a gas of "places" (holes). The state of the ensemble of this gas of holes is uniquely determined by the ensemble of the electrons, $p_{i,\mathrm{hole}} = 1 - p_i$. If, for two subsystems of the electrons, $p_{ij} = q_i r_j$, then, for the corresponding ensembles of holes, we have $p_{ij,\mathrm{hole}} = 1 - q_i r_j$, and the corresponding product for the subsystems reads,

$$(1-q_i)(1-r_i) = 1 - q_i - r_j + q_i r_j \neq p_{ij,\mathrm{hole}}.$$

Therefore, subsystems of the electrons are dependent even for the multiplicative $p_{ij} = q_i r_j$.

One can term this the effect of hidden components. It should be stressed that we speak, in fact, about an incomplete description (both the ensembles are uniquely related to each other), namely, that there are hidden components whose entropy has to be taken into account.

A different example, without any reference to quantum effects, is the entropy of monolayers on the surface of a solid (see, e. g. [20]). In the simplest case, the entropy density, up to constant factors and constants, has the form,

$$S = -c_{AZ} \ln(c_{AZ}/c^*_{AZ}) - c_Z \ln(c_Z/c^*_Z), \tag{29}$$

where A denotes molecules of the gas, Z is the vacant position on the surface (adsorbing center), AZ is the adsorbed molecule, c denotes corresponding surface concentrations, * denotes equilibrium concentrations, and since $c_Z + c_{AZ} = \mathrm{const}$ (the number of places per unit area is conserved) Eq. (29) is again the Fermi-Dirac entropy, obtained without any relation to quantum effects.

Thus, to sum up, the simplest known version of violating the additivity implies existence of subsystems of "locations", "holes", "ghosts" and like. These subsystems occupy the same states as the "observed" system, with probabilities,

$$\begin{aligned} q_i &= 1 - ap_i, \ a \in [0,1], \ \text{or} \\ q_i &= (1-a) + ap_i \end{aligned} \tag{30}$$

(We have stressed two possible cases, with a positive and with a negative constraint.) There might be several such hidden subsystems, and thus

$$S = S(\boldsymbol{p}) + \sum_j a_j S_j(\boldsymbol{q}^{(j)}), \tag{31}$$

where j is the label of the hidden subsystem, and $a_j > 0$. What the hidden subsystems could be? For example, they can describe various nonideal effects like excluded volume in various spaces (not obligatory in the physical R^3, as in the example of adsorbing centers). Other interpretations are probably possible. Here we do not consider any specific examples. Rather, we want to emphasize remarkable approximation possibilities provided by expression (31) when the entropies S_α are used. Indeed, already for just one hidden subsystem we have four fitting parameters (two coefficients in front of the Burg component for the system and for the ghost, one coefficient a in the constraint (30), and $1-a$ for the ghost). Because approximations to the experimental data obtained by the maximum entropy principle under certain constraints, whereas the choice of these coefficients is yet another optimization problem, it is not difficult, in principle, to organize a procedure of choosing the parameters (learning or fitting) in such a way as it is done in neural networks based on the error back propagation algorithm. Each time, an intriguing question will be arising, as to how many ghosts are needed for a description with a given accuracy, and what is the physical interpretation of those.

VI. HIDDEN SUBSYSTEMS CHANGING KINETICS

So far, all our considerations of the entropies, including either the Tsallis entropy, or the family S_α, as well as of the entropy of the ghosts (31) have left a nice option that they allow to do nothing about the master equation. Namely, all these entropies can be used to describe all kinds of incompletely known or restricted equilibria, or for constructing (generalized) canonical ensembles of dynamically conserved or quasi-conserved quantities. If the probability evolves in time according to the master equation, all these entropy functions behave correctly, that is, they monotonically increase with the time. This fact is well known, and it was reviewed above in section II.

In other words, as long as the hidden subsystem is described by the same set of states, as the observed one, no restrictions on the Markov kinetic equation arise. However, if more freedom is allowed in the choice of the entropy, kinetics has to be modified. Indeed, for Markov chain satisfying the detail balance condition (7), the natural condition which defines the equilibrium of the transition $p_i \rightleftharpoons p_j$ can be written,

$$\frac{\partial S}{\partial p_i} = \frac{\partial S}{\partial p_j}. \tag{32}$$

For the entropy function of the form $S = -H_h$, where H_h is given by equation (5), this writes,

$$h'(p_i/p_i^{\rm eq}) = h'(p_j/p_j^{\rm eq}), \tag{33}$$

and, because of strict monotonicity, this results in the definition of the equilibrium, $p_i/p_i^{\rm eq} = p_j/p_j^{\rm eq}$. Master equation (3) with the detail balance condition can be rewritten in such a way as to make it apparently consistent with the latter result: We denote,

$$w_{ij} = q_{ij}p_j^{\rm eq},$$

then master equation (3) can be rewritten,

$$\dot{p}_i = \sum_{j=1}^{N} w_{ij}[(p_j/p_j^{\rm eq}) - (p_i/p_i^{\rm eq})]. \tag{34}$$

However, if the entropy of the hidden subsystem does not have the form $S_h = -H_h$, with H_h given by Eq. (5), for example, if it includes terms like,

$$(a_0 + \sum_{i=1}^{N} a_i p_i)\ln(a_0 + \sum_{i=1}^{N} a_i p_i),$$

then condition (32) results in a more complicated equation, which, unlike Eq. (33), brings in the dependence on all the components of the vector $\boldsymbol{p}$. In that case, a model kinetic equation, more general than the master equations can be addressed.

There exist a universal way of constructing kinetic equations in a way consistent with the given entropy. Let us introduce notation, $\mu_i = -\partial S/\partial p_i$, and let $\Psi(x)$ be a monotonically increasing function. Then we define the rate of transitions $p_i \to p_j$ as

$$w_{ij}(\boldsymbol{p})\Psi(\mu_i), \tag{35}$$

where $w_{ij} = w_{ji}$, $w_{ij} \geq 0$ is a symmetric matrix with nonnegative entries (which might also be functions of the probability distribution $\boldsymbol{p}$). Given the rates (35), kinetic equation takes the form:

$$\dot{p}_i = \sum_{j} w_{ij}(\boldsymbol{p})(\Psi(\mu_j) - \Psi(\mu_i)). \tag{36}$$

Equation (36) is a generalization of the Marcelin – De Donder kinetic formalism (see, for instance, [11], [20], [21], [4]).

Equation (36) becomes especially natural to use if the entropy of the system has the form,

$$S = S_h + \tilde{S}, \tag{37}$$

where the part $S_h = -H_h$ has standard form (5) for some convex function h while $\tilde{S}$ is the entropy of the (part of the) hidden subsystem which does not have such form. Then we put, $\Psi(x) = [-h']^{-1}(x)$, that is, Ψ is the inverse of the derivative $-h'$. With this, Eq. (36) becomes:

$$\dot{p}_i = \sum_{j=1}^{N} w_{ij} \left([-h']^{-1}(h'(p_j/p_j^{\mathrm{eq}}) - \partial\tilde{S}/\partial p_j) - [-h']^{-1}(h'(p_i/p_i^{\mathrm{eq}}) - \partial\tilde{S}/\partial p_i) \right). \tag{38}$$

This is the minimal extension of the master equation: If the hidden system can be described with the same entropy ($\tilde{S} = 0$), equation (38) reduces to master equation. However, in any case, extensions (36) and (38) are consistent with the entropy increase in the kinetic processes.

VII. DISCUSSION

Once a classical statistical system is out of the thermodynamic limit, the exclusive character of the Boltzmann-Gibbs-Shannon entropy is lost, and classical ensembles are not equivalent anymore. Whereas using the microcanonical ensemble for any description of finite systems may be most appropriate, this route is very complicated from a computational standpoint. For that reason, seeking an entropic description of effects of finiteness seems a relevant option.

In this paper, we have demonstrated that that there exists a unique one-parametric family of entropy functions which are consistent with the additivity of the entropy under joining statistically independent subsystems. This family is essentially the convex combination of the Boltzmann-Gibbs-Shannon entropy and the Burg entropy. This family of entropy functions appears in a natural way as a distinguished (by the additivity requirement) subset of the family of Lyapunov functions of master equation. It has been demonstrated that the nontrivial contribution from the Burg component results in broadening of the high-energy tail of the canonical distribution function. The functional form of the deviation, and, in particular, the appearance of the energy correlations indicates that the maximum entropy approach successively used recently in the context of Tsallis entropy may lead to similar results when the present entropy functions are used. Detailed study of this option is left for the future work.

Finiteness of the classical statistical system is one option which calls for non-classical entropies. A different (independent) option is the incompleteness of the description. This has been demonstrated by analyzing the classical example of Fermi-Dirac type of entropy, and a generalization in the form of "standard entropy for a multi-component mixture plus linear constraints" has been suggested. Finally, we have suggested a minimal modification of master equation consistent with the given entropy.

---

[1] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).

[2] http://tsallis.cat.cbpf.br/biblio.html.

[3] E. Vives and A. Planes, Phys. Rev. Lett. **88**, 020601 (2002).

[4] I. V. Karlin, M. Grmela, and A. N. Gorban, Phys. Rev. E **65**, 036128 (2002).

[5] D. Ruelle, *Statistical Mechanics. Rigorous Results* (Benjamin, New York, 1969).

[6] T. Koga, J. Chem. Phys. **22**, 1633 (1954).

[7] A. M. Kogan, J. Appl. Math. Mech. **29**, 130 (1965).

[8] R. M. Lewis, J. Math. Phys. **8**, 1448 (1967).

[9] A. N. Gorban and I. V. Karlin, Physica A **206**, 401 (1994).

[10] C. D. Levermore, J. Stat. Phys. **83**, 1021 (1996).

[11] A. N. Gorban, *Equilibrium Encircling. Equations of Chemical Kinetics and Their Thermodynamic Analysis* (Nauka, Novosibirsk, 1984).

[12] A. Rényi, *Probability Theory* (North-Holland, Amsterdam, 1970).

[13] A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).

[14] J. P. Burg, paper presented at the 37th meeting of the Society of Exploration Geophysicists, Oklahoma City, 1967.

[15] J. P. Burg, Geophysics **37**, 375 (1972).

[16] I. V. Karlin, A. Ferrante, and H. C. Öttinger, Europhys. Lett. **47**, 182 (1999).

[17] B. M. Boghosian, J. Yepez, C. P. Coveney, and A. Wagner, Proc. Roy. Soc. Lond. A MAT

457, 717 (2001).

[18] C. Beck, Europhys. Lett. **57**, 329 (2002).

[19] M. Kac, *Probability and Related Topics in Physical Sciences* (Interscience, NY, 1957).

[20] G. S. Yablonskii, A. N. Gorban, V. I. Bykov, and V. I. Elokhin, *Kinetic Models of Catalytic Reactions. Comprehensive Chemical Kinetics, V. 32, ed. by G G. Compton* (Elsevier, Amsterdam, 1991).

[21] M. Grmela, Can. J. Phys. **59**, 698 (1981).