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, Sonneggstrasse 3, ML J 19, CH-8092 Zürich, Switzerland
(Received 27 March 2002; published 16 January 2003)

We derive a one-parametric family of entropy functions that respect the additivity condition, and which describe effects of finiteness of statistical systems, in particular, distribution functions with long tails. This one-parametric family is different from the Tsallis entropies, and is a convex combination of the Boltzmann-Gibbs-Shannon entropy and the entropy function proposed by Burg. An example of how longer tails are described within the present approach is worked out for the canonical ensemble. We also discuss a possible origin of a hidden statistical dependence, and give explicit recipes on how to construct corresponding generalizations of the master equation.

DOI: 10.1103/PhysRevE.67.016104 PACS number(s): 05.70.Ln, 05.20.Dd

I. INTRODUCTION

The past several years have witnessed a burst of interest in nonextensive statistical mechanics, a topic that finds increasingly more applications due to the concept of the 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) extends 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 results associated with the 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. With this, one can address long (nonexponential) tails of probability distributions. The characteristic feature of Tsallis' entropy is its nonextensivity for $q\neq 1$. If the system is composed of two statistically independent subsystems then Tsallis' entropy of this system is not equal to the sum of Tsallis' entropies of the subsystems. Since Tsallis' entropy is postulated rather than derived, this point remains open to discussion [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 that can model effects of finiteness.

We first remark that real-world systems, to which statistical mechanics is applied, are finite and, though they consist of a large number of subsystems, the natural logarithm of this number is not that big after all, it is not larger than 100 and is often less than 20 (since we address questions related to entropies, one should estimate the magnitude of the logarithm, in the first place). Extensivity in the true sense of this notion, theorems of equivalence of the microcanonic and the canonic ensembles [5] and the like, are 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 ν of such independent and noninteracting subsystems, which are similar in all their observable properties to the larger system, that plays the role of the parameter whose value tells us how close the system is to the thermodynamic limit.

One realizes that ν is finite when it is needed to cut off the tails of the distribution functions with divergent averages (with this, one restores the argument about the incomplete extensivity). This is a well known fact, for example, in the case of the classical Boltzmann equation. The maximum entropy solution to 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 [6–10]. 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 interactions 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 accurate universal description in the sense just mentioned cannot be ruled out *a priori*. For that reason, a search for nonclassical entropies for a possible description of nonextensive systems seems to be motivated.

*Electronic address: gorban@icm.krasn.ru
†Electronic address: ikarlin@ifp.mat.ethz.ch

The structure of this paper is as follows. In Sec. II, we review, for the sake of completeness, the theory of Lyapunov functions of the master equation. In Sec. III we derive the family of the (almost) additive entropies from the condition of additivity for statistically independent systems. In Sec. 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 Sec. V we discuss a different scenario, how an apparent statistical dependence may occur when the description of the system is incomplete. We also develop a natural generalization of the master equation for the situations with an incomplete description in Sec. VI. Finally, results are discussed briefly in Sec. VII.

II. LYAPUNOV FUNCTIONS OF MASTER EQUATION

We begin 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 occupy only these states. The probability distribution at time $t\geqslant 0$ is given by the N-component vector $\boldsymbol{p}$ with components $p_i(t)$, where $p_i\geqslant 0$, and $\Sigma_{i=1}^N p_i=1$. Equation for $\boldsymbol{p}$,

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

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

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

The graph of transitions is put in correspondence to each Markov chain by drawing the oriented link from the node E_i to the node E_k if $q_{ki}>0$. The important subclass of Markov chains is characterized by directional connectivity. The graph of transitions is called directionally connected if there is a path built of oriented links from each node to any other node. Then the following ergodic theorem is valid. Let the transitions 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 $\boldsymbol{p}(0)$, the solution $\boldsymbol{p}(t)$ to Eq. (3) tends to $\boldsymbol{p}^{\mathrm{eq}}$ at $t\to\infty$.

In the sequel we consider only a Markov chain which satisfies the ergodic theorem. Let us assume that the stationary state $\boldsymbol{p}^{\mathrm{eq}}$ is known, and let $h(x)$ be a convex twice differentiable function of one variable $x\in[0,\infty]$. Any function h with these properties defines a convex Lyapunov function H_h of the Markov chain (3) by the following rule:

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

The time derivative of the function H_h (5) due to Eq. (3) is nonpositive,

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

where the prime denotes derivative with respect to argument. The equality sign is reached only in the stationary state. The stationary state $\boldsymbol{p}^{\mathrm{eq}}$ is called the state of detail balance, if it satisfies

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

Markov chains with detail balance are colloquially termed master equations. In this case, the time derivative of the Lyapunov function becomes especially simple,

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

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 functions is again a convex function, the obvious construction that enables one to construct other Lyapunov functions from given representatives of the family (5) is this: If $h_1,\dots,h_k$ are convex functions, and if $\alpha_1,\dots,\alpha_k$ are non-negative and satisfy $\Sigma_{m=1}^k\alpha_m=1$, then

$$H_{\alpha_1h_1,\dots,\alpha_kh_k}=\sum_{m=1}^k\sum_{i=1}^N p_i^{\mathrm{eq}}\alpha_m h_m(p_i/p_i^{\mathrm{eq}}), \tag{9}$$

is also the Lyapunov function of the Markov chain. It should be stressed that the set (9) does not extend 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 using 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 (see the following 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 $\boldsymbol{q}$, $p_i \geqslant 0$, where $q_j \geqslant 0$, $\Sigma_i p_i = 1$, $\Sigma_j p_j = 1$. First, we will consider the case of the equipartition at the equilibrium, in order to simplify notation (a generalization to the arbitrary case is straightforward, see remark 5 below). Specifically, we assume that the equilibrium states of both the systems are equipartitions with probability vectors $\boldsymbol{p}^{\mathrm{eq}}$ and $\boldsymbol{q}^{\mathrm{eq}}$, where $p_i^{\mathrm{eq}} = 1/P$, $q_i^{\mathrm{eq}} = 1/Q$, and where P and Q are the numbers of the states in each of the the systems.

Since the systems are independent, the joint system is characterized by the joint probability vector $\boldsymbol{p}\cdot\boldsymbol{q}$. The equilibrium of the joint system is again the equipartition, $(\boldsymbol{p}\cdot\boldsymbol{q})^{\mathrm{eq}} = \boldsymbol{p}^{\mathrm{eq}}\cdot\boldsymbol{q}^{\mathrm{eq}}$, 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,

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

This functional equation has two special solutions that correspond 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. The function H_1 corresponds to the classical (additive) Boltzmann-Gibbs-Shannon entropy, thus, we demonstrate here the additivity of H_2 only. Indeed,

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

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

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

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 nonclassical contribution due to H_2 becomes significant only if the system is not too close to the thermodynamic limit. Only the classical Boltzmann-Gibbs-Shannon contribution survives in the thermodynamic limit.

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 using either $\boldsymbol{p}$ or $\boldsymbol{p}^{\mathrm{eq}}$. The latter is the distinguished probability distribution which, same as $\boldsymbol{p}$, is multiplicative with respect to joining the statistically independent subsystems. Relevance of the master equation, and hence of the kinetic rather than of the static picture, to our derivation of the one-parametric family (11) is clear. 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 that 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.

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 the following 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 $0 \leqslant p_i \leqslant 1$, and its replacement by continuity on semiopen intervals, $0 < p_i \leqslant 1$.

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

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

was first considered by Burg in the context of applications of information theory to geophysical problems [14,15]. Recently, the Burg entropy (12) was used to construct examples 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) was mentioned as just the solution to the additivity condition, its relevance to describing effects of finiteness in statistical systems was not duly discussed.

Remark 5. If the equilibrium $\boldsymbol{p}^{\mathrm{eq}}$ of the Markov chain differs from the equipartition but remains 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 describe the long tails for $\alpha \neq 1$. In the context of discrete system of states, the long tail has to be understood as a broadening of the distribution functions.

Since we are going to study the case of small α in this section, the factor $1/N$ in front of the second term in Eq. (11) will be omitted in order to simplify notation, and we consider the one-parametric set of entropy functions,

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

We shall consider first the microcanonic ensemble, that is, the minimizer of H_α under the constraint of fixed normalization, $\Sigma_{i=1}^{N} p_i = 1$. 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 $E_i \geqslant 0$, and find the minimum of H_α (14) under the constraints,

$$\sum_{i=1}^{N} p_i = 1, \quad (15)$$

$$\sum_{i=1}^{N} E_i p_i = U. \quad (16)$$

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

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

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 some algebra, we find for $\alpha \ll 1$, $\alpha > 0$,

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

where

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

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 $\beta^{(0)}(U)$ in terms of U], and

$$V = \sum_{i=1}^{N} E_i, \quad (20)$$

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

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

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

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, $dp_i^{(\alpha)}/d\alpha|_{\alpha=0}$? The canonical distribution (19) decays when E_i exceeds 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 answer this question, we obtain in Eq. (17),

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

where the factor A_i is

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

Factor A_i amplifies populations of the states that 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,

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

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, $E_i = E + \delta_i$, $\Sigma_{i=1}^{N} \delta_i = 0$, 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

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

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

$$\delta_i \geqslant 2/\beta_E^{(0)}. \quad (27)$$

In other words, raising of the populations of the higher-energy levels is explicitly demonstrated by 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 an explicit example that taking into account the Burg component in the one-parametric family (11) indeed is able to describe the broadening of the canonical distribution function. Appearance of the energy levels correlation in the above formula (18) remarkably resembles recent results of applications of the Tsallis entropy to the fittings of experimental data in turbulence (see Ref. [18] and references cited therein). Generalizations to the quasiequilibrium situation with more constraints is straightforward. In the case of the continuous states, the long tail feature of the corresponding distribution functions becomes even more transparent. 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, $P(x) \sim (\lambda + \beta x^2)^{-1}$ (the well-known Cauchy distribution). For it, algebraic decay at infinity precludes existence of the variance, and a cutoff is required.

In order to conclude, in Secs. III and IV we have demonstrated that for systems out of the thermodynamic limit there exists the universal (that is, independent of details of interactions) one-parametric family of additive entropy functions that are able to describe the same long tail effects as Tsallis' entropy. The one-parametric family of additive functions (11) or (14) is unique in the sense that any other convex function of the probability which satisfies simultaneously the additivity and the trace-form requirements is obtained from Eq. (11) or Eq. (14) by adding a constant or multiplication by a constant. In the remainder of this paper we shall discuss a different issue of how the additivity of the entropy can *apparently* be violated if the description of the system is incomplete.

V. NONADDITIVITY AND INCOMPLETE DESCRIPTION

In the recent literature, it is sometimes argued that the entropy is not additive under joining the statistically independent subsystems because, *in reality*, these subsystems are not independent. For example, "the concept of *independent* subsystems does not make any sense, since all subsystems are interacting" [18]. Possible physical agents such as long-range forces that could lead to such a situation are occasionally mentioned, though, to the best of our knowledge, it has not yet been demonstrated for any realistic system. Moreover, caution is needed in rejecting the "concept of independent subsystems" because this leads to a confrontation with the traditional axiomatic (Kolmogorov's) probability theory that is strongly based on this concept [19] (in the worst case, one should abandon the concept of independent trials, which is at the very heart of the definition of the probabilities).

This all leads to one question. If the probability distribution over pairs of states p_{ij} factors into the product of the distributions, $p_{ij}=q_i r_j$, but the subsystems are dependent, then where is this dependence hidden? 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

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

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 the 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,

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

Thus, subsystems of the electrons are dependent even for the multiplicative $p_{ij}=q_i r_j$. It should be stressed that, in fact, we speak of 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 is the entropy of monolayers on a solid surface (see, e.g., Ref. [20]). In the simplest case, the entropy density, up to constant factors and constants, has the form,

$$S = -c_{AZ}\ln(c_{AZ}/c_{AZ}^{\mathrm{eq}}) - c_Z \ln(c_Z/c_Z^{\mathrm{eq}}), \qquad (29)$$

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. Since $c_Z + c_{AZ} = \mathrm{const}$ (the number of places per unit area is conserved), the entropy Eq. (29) is again of the Fermi-Dirac form which is now obtained without any relation to quantum effects.

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

$$q_i = 1 - a p_i, \quad a \in [0,1], \quad \text{or}$$

$$q_i = (1-a) + a p_i. \qquad (30)$$

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

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

where j is the label of the hidden subsystem, and $a_j > 0$. Hidden subsystems can describe effects such as excluded volume in various spaces (not obligatory in the physical R^3, as in the example with the adsorbing centers). Other interpretations are probably possible. Here we do not consider any specific example. Rather, we want to emphasize the remarkable approximation possibilities provided by the 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 are obtained by the maximum entropy principle under certain constraints, whereas the choice of these coefficients is yet another optimization problem, it is not difficult 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.

VI. HIDDEN SUBSYSTEMS CHANGING KINETICS

All the entropies discussed above, including either Tsallis' entropy or the family S_α, can be used to describe incompletely known or restricted equilibria, for constructing (generalized) canonical ensembles of dynamically conserved or quasiconserved quantities. If the probability evolves in time according to the master equation, all these entropy functions behave equally correctly, that is, they monotonically increase with the time (see Sec. II). In other words, as long as the hidden subsystem is described by the same set of states, as the observed one, no restrictions arise on the Markov kinetic equation. This situation becomes different if more freedom is allowed in the choice of the entropy. Before describing the corresponding generalization of the master equation, it is instructive to consider again the standard case.

For a Markov chain consistent with the detail balance condition (7), the *natural condition* that defines the equilibrium of the transition $p_i \rightleftharpoons p_j$ can be written as follows:

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

For the entropy function of the form $S=-H_h$, where H_h is given by Eq. (5), the latter equation gives,

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

Furthermore, thanks to the strict monotonicity of the derivative h', this results in the usual definition of the equilibrium, $p_i/p_i^{\mathrm{eq}}=p_j/p_j^{\mathrm{eq}}$. Master equation (3) with the detail balance condition can be written in such a way as to make it explicitly consistent with the latter result. Introducing notation, $w_{ij}=q_{ij}p_j^{\mathrm{eq}}$, master equation (3) can be cast into the following form:

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

Though simple, the above derivation should be appreciated because the natural definition of the equilibrium (32) results in the equilibrium between pairs of states only. However, if the entropy of the hidden subsystem is not of the form $S_h=-H_h$, with H_h given by Eq. (5), for example, if it includes terms like

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

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

Let us introduce notation, $\mu_i=-\partial S/\partial p_i$, and let $\Psi(x)$ be a monotonically increasing function. We define the rate of transitions $p_i \to p_j$ as

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

where $w_{ij}=w_{ji}$, $w_{ij}\geqslant 0$ is a symmetric matrix with nonnegative matrix elements (matrix elements are allowed to depend on the probability distribution $\boldsymbol{p}$). Given the rates (35), the generalized 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, e.g., Refs. [11,20,21,4]).

Equation (36) is 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 part of entropy function of a different form. Then we put $\Psi(x)=[-h']^{-1}(x)$, that is, Ψ is the inverse of the derivative $-h'$. With this, Eq. (36) becomes

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

This is the minimal extension of the master equation. If the hidden system can be described with the same entropy ($\tilde{S}=0$), then Eq. (38) reduces to the master equation. By construction, both Eq. (36) and Eq. (38) are consistent with the entropy increase in the kinetic processes, and therefore can be used in modeling the kinetic processes. Other kinetic representations specializing to the Tsallis entropy can be found in the literature (see, e.g., Refs. [22,23]).

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 the computational standpoint. For that reason, seeking the entropic description of effects of finiteness is a relevant option.

In this paper, we have demonstrated that there exists the unique one-parametric family of entropy functions that 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 of the Burg entropy. This family of entropy functions appears in a natural way as the distinguished (by the additivity requirement) subset of the family of Lyapunov functions of the master equation. It has been demonstrated that the nontrivial contribution from the Burg component results in a 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 the 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 classical statistical systems is one option that calls for nonclassical entropies. A different (independent) option is the incompleteness of the description. This has been demonstrated by analyzing the classical example of the Fermi-Dirac type of entropy, and a generalization in the form of ‘‘standard entropy for a multicomponent mixture plus linear constraints’’ has been suggested. Finally, we have suggested a modification of the 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 (unpublished).
[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. R. Soc. London, Ser. A **457**, 717 (2001).
[18] C. Beck, Europhys. Lett. **57**, 329 (2002).
[19] M. Kac, *Probability and Related Topics in Physical Sciences* (Interscience, New York, 1957).
[20] G. S. Yablonskii, A. N. Gorban, V. I. Bykov, and V. I. Elokhin, in *Kinetic Models of Catalytic Reactions. Comprehensive Chemical Kinetics*, edited by G G. Compton (Elsevier, Amsterdam, 1991), Vol. 32.
[21] M. Grmela, Can. J. Phys. **59**, 698 (1981).
[22] M. Buiatti, P. Grigolini, and A. Montagnini, Phys. Rev. Lett. **82**, 3383 (1999).
[23] M. Bologna, C. Tsallis, and P. Grigolini, Phys. Rev. E **62**, 2213 (2000).