Statistical physics and information theory for systems with local constraints

Zhang, Q.

Chapter 1

Introduction

Statistical physics is built to explain the macroscopic properties in physical systems from the probabilistic examination of underlying microscopic configurations [1]. As the initial research of statistical physics is focused on thermodynamic systems, most traditional examples in statistical physics are under global constraints, such as the fixed total energy and the fixed number of particles. However, recent research on complex systems shows that local constraints, which is implied by the heterogeneous spatial and temporal dependencies, general existed in natural systems [2–4]. Unlike the global constraint, which works on all units in the system with the same influence, the local constraints work on the different groups of units in the system with different influences. The local constraints are not only a new model, which can be used to describe the non-physical systems with heterogeneous dependencies. It also brings several new phenomenons for statistical physics. In particular, ensemble equivalence is broken in networks with fixed degree sequences by the extensive local constraints [5–7]. This breaking of ensemble equivalence exists in the whole parameter space of networks and even without the appearance of phase transitions, which is the essential condition of ensemble nonequivalence in traditional statistical physics. Ensemble equivalence as a basic assumption in statistical physics has been widely explored both practically and theoretically [8]. This breaking of ensemble equivalence in systems with local constraints will affect many fundamental assumptions and calculations in statistical physics. Therefore, further exploration of statistical physics needs a general theory to describe the system with local constraints and the possible appearance of ensemble nonequivalence.

Moreover, local constraints in complex systems also bring new problems to the information theory. For instance, the activities of neurons in nervous systems with heterogeneous spatial correlations give an information source with numerous interacting units [9]. The temporal dependent fluctuation in financial systems breaks the identical independent assumption of the information sources in the classical information theory [3]. The behaviour of those new information sources is impossible to be described by the random variables with finite outcomes as in classical information theory. To describe them, we need statistical ensembles with local constraints. Particularly, the possible appearance of ensemble nonequivalence that is caused by the extensive local constraints may even break the asymptotic equipartition property in the information theory and affect the information-theoretical bounds. Hence, a non-trivial generalization of the information theory for the system with local constraints is needed to study the information transmission and storage in complex systems.

This thesis is an exploration of systems with local constraints. The first part introduces two physical models with local constraints and studies the breaking of ensemble equivalence of it. In the second part, these physical models are used to describe information sources and sequences with heterogeneous dependencies to find new information-theoretical bounds and the influence of ensemble nonequivalence in it.

The introduction has four sections. I will introduce the definition of statistical ensembles in section1.1. The basic conception of ensemble nonequivalence is introduced in section1.2. The classical Shannon information theory is introduced in section1.3. Section1.4 is the outline of this thesis.

1.1 Statistical ensembles

In the research on thermodynamic systems, to describe the behaviour of numerous random interacted particles is a difficult task, e.g., a glass of water can easily contain 10^{23} molecules. Classical mechanics can be used to describe the collision between the finite number of water molecules, but it is impossible to understand the process of collision with 10^{23} units analytically. This insurmountable problem prompts physicists to build new mechanics to describe the macroscopic behaviour of the system and still based on the physical law already know from the microscopic view. This requirement leads to the birth of statistical physics in the 19th century [1, 10].

Generally, the founding of statistical physics is credited to three physicists. Ludwig Boltzmann gives the fundamental definition of entropy by the collection of microstates [10]. James Clerk Maxwell applies the model of the probability distribution into the description of those microstates [11]. Josiah Willard Gibbs defines statistical ensembles and explains the laws of thermodynamic as the statistical properties of the ensembles [1]. This pattern also can be found later in the establishment of information theory. The initial study focuses on the analytical configurations, but the theory is established on probabilistic explanation.

The framework of statistical physics can be explained as the application of statistical methods and probability theory to large assemblies of microscopic entities and then using the mathematical tools for dealing with large populations and approximations to connect the behaviour of the microscopic entities with the macroscopic property. There are three basic postulates in statistical physics. The ergodic hypothesis shows that each state with the same energy has the equiprobable probability to appears in the system over long periods. The principle of indifference shows that we can only assign equal probabilities to each state when there is insufficient information to describe states in this system. The maximum information entropy presents that the correct probability distribution of the states in the system with limited information should maximum the Gibbs entropy (Information entropy) of it [1]. In the past decades, the principle in statistical physics is not only applied in the physics system but also generalized in chemistry, biology, economics, and even social science [12–14], to build the relationship between the microscopic entries and the macroscopic property of them.

The building of the statistical ensembles needs states in thermodynamic systems to remain 'static'. This static is that the macroscopic observable variables in the system keep stable under the unpredictable internal particles' motion. These observable variables can be the total energy, the temperature, and the pressure [1]. The difference between macroscopic variables allows us to describe the systems by different ensembles.

- *Microcanonical ensemble* is used to describe the system with a fixed total energy and particle numbers. In this isolated system, the energy change between the system and the outside environment is forbidden. States in this ensemble have the same probability of appearing in the system. The value of the probability is decided by the total number of microscopic configurations in this system.

- The *Canonical ensemble* is used to describe the system with fixed particle numbers and temperature. This system contacts a heat bath, which has a precise temperature. There is an energy change between this system and the heat bath. The probability of each state appears in this system is decided by their total energy under the fixed temperature.

- The *Grand canonical ensemble* is used to describe the open system. Both the total energy and the number of particles are not fixed. Systems described by this ensemble will change energy and particles with the outside environment. The probability of each state in this ensemble is decided by their total energy and their total particles.

According to the energy isolation, the microcanonical ensemble and the canonical ensemble are under two different constraints. The microcanonical is under 'hard' constraints, as all states should have the same total energy. But the canonical ensemble is under 'soft' constraints, the total energy of each state in this ensemble is not the same, but the average value should fix.

As the way to describe the interactions are decided by the property of constraints of the statistical ensemble, the soft or hard constraints in the systems not only decide which statistical ensemble is suitable for their descriptions but also affect the way to describe the inner interactions [15, 16]

The traditional statistical ensemble is under global constraints like the total energy, the fixed temperature or the total number of particles in the system. But the information sources and the process of signal generating of the complex systems are all under heterogeneous interactions. It means the constraints in the information source or the signal generation need to be localized to describe those heterogeneous interactions. The localized constraint is not a new concept, it has been studied in networks theory, where the nodes in the networks always have different degrees, and this heterogeneous degree distribution will affect the dynamic and structure of the networks [16–20].

However, systems that need to model in the new information theory is more general than the binary network with fixed degree sequences. The interactions among units in information sources may have different degrees. Thus, using the weighted network to model those heterogeneous dependencies is more reasonable than the binary networks. Furthermore, when the signal generated by the finite variables under temporal dependences in long periods needs to record, the data structure we need is not the adjacency matrix. This data structure is close to an $m \times n$ matrix, where m represents the number of finite variables and n is the increasing length. Therefore, to describe the systems mentioned above, new models with local constraints are needed.

1.2 Ensemble (non)equivalence

Normally, the three ensemble descriptions are unavoidable different when the system has a finite size. There are fluctuations in the macroscopic properties of states in the canonical and grand canonical ensemble. But when the system has numerous particles (in the thermodynamic limit), all the three ensembles tend to give an identical description. The fluctuation of macroscopic properties will vanish. This is a basic assumption in the tradition statistical physics, which is named as ensemble equivalence [1, 8, 21].

Specifically, in the microcanonical ensemble and the canonical ensemble, this equivalence appears in the systems with the same total number of particles but under 'soft' or 'hard' constraints (according to the energy isolation). Since the canonical ensemble is mathematically easy to calculate, the presence of the ensemble equivalence means that the replacing of the microcanonical ensemble with the canonical ensemble in the application of statistical physics is allowed [8]. The ensemble equivalence has three forms,

- *Thermodynamically equivalence*: When the entropy of the microcanonical ensemble and the free energy of the canonical ensemble is one-to-one related under the Legendre transform, the two ensembles are believed under thermodynamical equivalence.

- *Macrostate equivalence*: The macrostate equivalence is the equilibrium values of the macrostate predicted by the microcanonical ensemble is the same as the one of the canonical ensemble in the thermodynamic limit.

- *Measure level equivalence*: when the probability distribution of states in the canonical ensemble Gibbs distribution)converges to the probability distribution of the states in the microcanonical ensemble that is defined by the Boltzmann's equiprobability postulate, we believe the two ensembles are under measure level equivalence.

The presence of ensemble equivalence in the thermodynamic limit shows that although the specific action in the canonical ensemble varies from one microstate to another, most of the microstates are still roughly equiprobable.

However, this ensemble equivalence does not always hold. Recent researches on fluid turbulence [22, 23], star formation [24, 25] and networks [26] show that the ensemble equivalence will break at the critical point on the boundary of phase transition.

This ensemble nonequivalence is caused by the nonadditivity in the system with long-range interaction [13, 8]. More recently, in networks with fixed degree sequences, the breaking of ensemble equivalence even has been found in the whole parameter space under the complete absence of phase transitions [5, 7, 26]. It means the ensemble equivalence is not only a critical phenomenon but also an intrinsic property of the system with extensive local constraints [27, 28]. Therefore, there will always have a non-neglected difference between different ensembles in the ensemble nonequivalent systems.

The three forms of equivalence have three coincide ensemble nonequivalences, and those three forms already proved equivalent in [8]. For instance, the thermodynamical nonequivalence is the function of microcanonical entropy, not one-to-one relates to the *Legendre transform* of the canonical ensemble. It can be mapped by the difference between the prediction of macroscopic property for different ensembles. And it also can be detected by the difference between the probability distribution of the states in different ensembles [8].

The *measure level ensemble equivalence* is based on the difference between the probability distributions of different ensembles [8]. States in the microcanonical ensemble with the same total energy all belong to the conjugate canonical ensemble [15]. The measure-level ensemble equivalence between the two ensembles is that the probability distribution of the states in the canonical ensemble converges to the probability distribution of the states in the microcanonical ensemble. Therefore, under measure level ensemble nonequivalence, the difference between two probability distributions will not vanish, and it can be quantified by the relative entropy between the microcanonical and canonical ensemble as

$$S[P_{\text{mic}}||P_{\text{can}}] = \sum_{x \in \S} P_{\text{mic}}(x) \ln \frac{P_{\text{mic}}(x)}{P_{\text{can}}(x)}, \tag{1.1}$$

where § is the collection of all the possible microscopic configurations of the system with n particles in it. Symbol $P_{\text{mic}}(x)$ represents the probabilty of the state x in the microcanonical ensemble. Symbol $P_{\text{can}}(x)$ represents the probability of the state x in the canonical ensemble.

The indicator of the measure ensemble nonequivalence is the specific relative entropy density [8, 5], which is defined as the limit of the relative entropy rescaled by the number of particle in it

$$s_\infty = \lim_{n\to\infty} \frac{1}{n} S[P_{\text{mic}}||P_{\text{can}}]. \tag{1.2}$$

When the value of s_∞ is equal to 0, we believe the system is under measure level ensemble equivalence. When $s_\infty > 0$, the system is believed to be under measure level ensemble nonequivalence [8].

Ensemble nonequivalence appears in the boundary of phase transition in the system with long-range interactions has been well studied theoretically and experimentally [13, 8]. But we know little awareness about the ensemble nonequivalence in the system with local constraints. Compared to the one that only happens on the critical point of parameter space, the ensemble nonequivalence in the system with local constraints has a more general form. In the binary networks with degree sequences, the ensemble nonequivalence appears in the whole parameter space of it [5]. No matter how the degree distribution of the network change, the probability distribution of the networks' configuration in the microcanonical always has a non-neglected difference between the canonical ensemble. This difference will affect the generalization of statistical ensembles to complex systems with heterogeneous interaction, especially in the information theory, as the information source and the information transmission are all described by the probability theory. Thus, a more fundamental study about the ensemble nonequivalence in the system with local constraints is needed.

Classical information theory has a close relationship with statistical physics. The increasing length of the sequences in information theory is the same as the extension of the system's size in statistical physics. The identical probability distribution of the information sources and the independent signal generating are all coincided with the definition of the microscopic behaviour of the particles in statistical physics under global constraints [29, 5]. Thus, the information-theoretical bounds as the macroscopic properties of the information storage and transmission are decided by the statistical properties of microscopic configurations. The ensemble nonequivalence that may appear in the information source with heterogeneous interaction or the information sequences with temporal dependence has a distinct influence on the information storage and transmission. In other words, the generalization of the information theory for the complex system should be based on the *ensemble nonequivalent* systems.

1.3 Information theory

The information theory is built to describe the information transmission and storage in communication system [29]. The birth of information theory in the first half of the twentieth century is believed to be stimulated by the dramatic development of electronic communication systems. At that time, the worldwide electronic telegraph network continuous works more than a half-century. The birth of the telephone and television has already completely changed the daily life of humans. But there is still do not have a quantifiable definition of the information [30]. The earliest attempt to quantify information can be traced to 1924 when Nyquist introduced a theoretical speed of information transmission depending on the change of voltage in the line [31]. In 1928, Hartley first used the word 'information' to describe the stuff flows in the communication and generalized the definition of Nyquist into the whole communication systems as the number of possible states of the symbols transmits in this system [32]. Afterwards, in 1948, based on the probability theory, Shannon gives the first quantifiable definition of information (information entropy) and the information-theoretical bounds of the information storage and transmission [33].

Shannon divided the communication system into three parts: the information source, the channel and the receiver. These three parts can be described independently by the probability theory [33]. Information generated by the information sources $\mathbf{x}$ can be quantified by the definition of information entropy $H(\mathbf{x})$, which is equal the expected value of each probability's logarithm as $H(\mathbf{x}) = -\sum_{x\in\mathbf{x}} p(x)\log p(x)$ [33].

This information theory solved the two main problems in the research on artificial communication systems at that time. One is the smallest space to store the information generated by the information source, and the other is the maximum speed of reliable information transmitted through a channel. Shannon found that the information generated by the information source $\mathbf{x}$ is carried in the sequences $\{x_1, x_2, \cdots, x_n\}$ to record the state of the sources in n times' activities. Therefore, the space to store the information generated by the information source is equivalent to the space to store those sequences. Simultaneously, Shannon also found that to store the information generated by the information sources only need to focus on the sequences in the typical set, as the sum of the probability of sequences in the typical set is close to 1. Therefore, the size of the typical set decides the size of the smallest space to store the information carried by those sequences. This typicality is based on the asymptotic equipartition property (AEP). According to the generalization of AEP into joint variables (source and receiver), Shannon also found the maximum speed of reliable information transmission through a channel is decided by the mutual information between the information source and receiver [33].

The rapidly developing information industry has demonstrated the effectiveness of classical information theory. And the application of it into biology, physics, and other disciplines also give new inspirations to solve the old problems or find new phenomenons in those systems [15, 34, 35]. Simultaneously, the applications also bring new issues back to the research of information theory. As in the systems like nervous systems with billions of neurons and social networks with billions of users [9, 36], the information sources are not a single variable but enormous interacted units. On the other hand, the process of signal generating by the information source is not independent. The probability of the information source to get different states is not identical, and it is constrained by the interactions from other units and the dependence from its past states [9]. Therefore, a new generalization of the information theory to deal with the heterogeneous dependencies in those systems is required.

To establish the new generalization information theory, we need to find new models to describe the information sources with enormous interacted units. And we also need to describe the process of single generating with heterogeneous dependence. The statistical ensembles from statistical physics are suitable for this duty [15], as they are build to describe the motion of particles with numerous interactions. The information sources with numerous interacted units need to be modelled by the statistical ensembles with local constraints. The temporal dependence in the process of the single generating that has broken the (identical independent distribution) i.i.d. assumption also can be modelled by the local constraints in the statistical ensemble. Therefore, this generalization is a combination of statistical physics and information theory for systems with local constraints.

1.4 Outline of this thesis

This thesis has two main parts: the first part is the study of systems with local constraints in statistical physics, which includes Chapter 2 and Chapter 3. The second part is the generalization of statistical ensembles with local constraints into information theory. It also has two chapters to introduce our contributions: chapter 4 and chapter 5.

In the second chapter, we focus on the weighted networks with core-periphery structures, which provides a possible model that has the phase transition and local constraint simultaneously, We find that relative fluctuation of constraints as a criterion to check the ensemble nonequivalence in traditional statistical physics vanish in the non-BEC phase while some of them do not disappear in the BEC phase. This result shows that fluctuations of constraints are sensitive to the phase transition. By contrast, the non-vanished relative entropy density for all positive temperatures shows that the extensive number of constraints breaks the ensemble equivalence. Only at zero temperature, where the effective number of constraints becomes finite, ensemble equivalence is broken by BEC in a subtle way. Therefore, in the presence of local constraints, the vanishing of relative fluctuations no longer guarantees ensemble equivalence.

In the third chapter, we extend the discussion of ensemble nonequivalence into a more general local constrained system, the $n \times m$ matrix ensemble $\mathbf{G}$. This matrix can be binary or weighted by setting the range of the value each entry archived. In this matrix ensemble $\mathbf{G}$, we can have the global constraint, one-side local constraints and two-side local constraints. In this general model, ensemble equivalence is still broken by the extensive number of constraints, both in the one-side local constraints (with n local constraints in it) and two-side local constraints (with $m+n$ local constraints in it). Surprisingly, when m is finite, the relative entropy of microcanonical and canonical ensemble has the same order as the canonical entropy, which is as strong as the one that appears in the boundary of phase transition caused by the non-activity. The result breaks the former empirical cognition that the ensemble nonequivalence should be 'strong' but 'restricted' as in the boundary of phase transition or 'general' but 'weak' as in the networks with degree sequences. Ensemble nonequivalence in the system with local constraints can be both 'general' and 'strong' when the units have a finite degree of freedom.

In the fourth chapter, the statistical ensembles with local constraints are used to describe the information source with numerous heterogeneous interacted units. We find the typical set of the microcanonical ensemble described information sources is a subset of the conjugate canonical ensemble described information sources. The classical information theory used to describe the multivariate information sources is a special case of the canonical ensemble description when the local constraints in the information sources are independent and identical. The microcanonical ensemble descriptions need less space to store the information generated by them. The extra sequences we counted in the canonical-typical set are those sequences, which have the same sum of *Hamiltonian* but different constraints comparing with the states in the microcanonical ensemble. The size of the extra space is decided by the degree of ensemble nonequivalence between the two ensembles. When the information sources are under strong ensemble nonequivalence, the space we can save by using the microcanonical ensemble has the same order as we cost in canonical ensemble description (in classical information theory). The information-theoretical bounds are directly affected by the ensemble nonequivalence in it.

In the fifth chapter, we focus on signal generating with coupled dependencies. The activity of variables in information sources is affected by the interactions between other variables and their historical behaviours. Each unit in the information source has two different dependencies: the spatial correlations come from the interaction between other variables in the source, and the temporal dependencies come from the historical sampling of itself. But all the information generated by these information sources is still carried by the information sequences, which are independent of each other. Thus, to find the limit of information storage and transmission under this situation, we need to use statistical ensembles with local constraints to describe the information sequences, not focus on the information source as in traditional information theory. We find that the breaking of ensemble equivalence in the signal generating is determined by the extensive spatial variational dependence among the variables in the information source, not the finite temporal dependencies of each variable itself. This result also explains why Shannon's classical information theory is so powerful. As in the classical information theory, there is only one variable or several independent variables in the information source, so there is no spatial dependence in this i.i.d. process. The sequences described by the classical information theory are under ensemble equivalence. The temporal dependence realized by the hard or soft constraints is equivalent to each other. The finite number of temporal dependence is not enough to break the ensemble equivalence between the canonical ensemble descriptions and the microcanonical ensemble descriptions. Thus, the signal generating process described by the classical information theory will approach the actual signal generating process, following the length increasing of the information sequences.

The last chapter gives the conclusion and some open problems about systems with local constraints.