©Copyright 2021
Yu-Chen Cheng

Asymptotic Behaviors and Perturbation Analysis of Stochastic Dynamics and Applications to Complex Systems

Yu-Chen Cheng

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Hong Qian, Chair

Krzysztof Burdzy

Matthew Lorig

Program Authorized to Offer Degree:

Applied Mathematics

University of Washington

Abstract

Asymptotic Behaviors and Perturbation Analysis of Stochastic Dynamics and Applications to Complex Systems

Yu-Chen Cheng

Chair of the Supervisory Committee:
Professor Hong Qian
Department of Applied Mathematics

The concept of hierarchical structures prevails among scientific points of view on complex systems. From one level to another in a hierarchical structure, with proper scales in both space and time, entire new laws emerge in the limits. Not only in natural science, but in formal science, mathematicians have widely applied this idea of scaling to prove several celebrated limit theorems. In the first part of the present work, new limit laws in conditional probability are shown. These new laws can be regarded as the mathematical foundation of Gibbisian canonical ensemble theory. Based on the new laws, the canonical ensemble theory can be generalized to strongly coupled heterogeneous systems. Another parallel canonical ensemble theory by Boltzmann is also discussed with applications to sample frequencies of the phenotype among a population of cells. In the second part, asymptotic behaviors of nonlinear dynamical systems under random perturbations have a full analysis. In particular, an underlying kinematic basis of the landscape of the dynamics emerges in the deterministic limit. These results provide a lens for helping us look through stochastic limit-cycle oscillations from different perspectives. In addition, entropy, entropy production, free energy dissipation of complex dynamics at the mesoscopic scale and the corresponding limiting behaviors at the macroscopic scale are depicted by the language of probability theory.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

ACKNOWLEDGMENTS

First and foremost I would like to thank my advisor, Professor Hong Qian for his support and mentoring over the years at the University of Washington. During my Ph.D. study, I always have his support whenever I need it. As a student from a very different background, I really appreciate that Professor Qian patiently taught me all the relevant knowledge to make this dissertation possible. His thought-provoking mentoring inspires me to pursue scientific truths. I am very fortunate to have a role model with great academic passion to guide my path in the future.

I would like to thank my thesis committee. I appreciate Professor Krzysztof Burdzy for introducing me to the beauty of probability theory. I was inspired a lot from the discussions with him in the philosophy of probability reading group. Professor Burdzy's insightful thought motivated me to find a solution to an important philosophical puzzle in this dissertation. I want to thank Professor Matthew Lorig. By seeing his dedication to the courses of applied stochastic analysis, I have learned how to be a good teacher, which is my dream in the future. I also enjoyed every talk with him when we occasionally met in the hallway of Lewis. I want to express my gratitude to my GSR, Professor Bertil Hille, for giving me very detailed comments on my general exam and helping me improve my work.

I particularly want to thank several other collaborators. I am thankful to Dr. Yizhe Zhu for helping me in my first pure mathematical project and making it possible. I have learned quite a lot of proof skills from him. I want to thank Professor Wenning Wang and Professor Zhiyue Lu for teaching me physics and chemistry in my first physics paper. I feel lucky to have had the opportunity to work with an exceptional group of collaborators: Lowell Thompson and Ying-Jen Yang. I have learned many new ideas and enjoyed the collaborative projects from them. I would also like to extend my gratitude to my graduate program advisor, Lauren Lederer: As an international student, it would have been impossible to complete my Ph.D. smoothly in a foreign country without her support.

Special thanks to the formal students in Hong Qian's group: Dr. Yue Wang and Dr. Felix Ye for sharing their valuable experiences in their Ph.D. careers. It was a great experience to study with them in the same group.

Among the friends I made during the Ph.D. years, I am very fortunate to work with them: Erin Angelini, Tyler Chen, Jeffrey Commons, Matthew Farrell, Kenan Li, Ben Liu, Kelsey Maass, Kelsey Marcinko, Sean Patrick Santos, Natalie Wellen, Cameron Wright, Xin Yang, Yang Zhou. With them from the beginning of this long journey, my Ph.D. study at University of Washington became one of the best periods of my life.

Finally, I want to share my accomplishment with my family. I want to thank my mom for her support. Every time I make a big decision, she is always by my side. And thank forever my wonderful wife, Shih-Hsuan, for her love and understanding. She is my best "counselor" whenever I face any difficulties. I am so lucky to have them in my life. Without them, I would not be able to accomplish this dissertation.

DEDICATION

This work is dedicated to Shih-Hsuan, my wonderful wife,
to my parents and grandparents for supporting me all the way.

Chapter 1

INTRODUCTION

Hierarchical structures are ubiquitous in complex systems. As the American theoretical physicist and Nobel laureate P. W. Anderson stated that [3] “At each level of complexity entirely new properties appear, and the understanding of the new behaviors requires research which I think is as fundamental in its nature as any other”. He enumerated a series of different levels of complex systems: few-body (particle) physics, many-body (statistical) physics, chemistry, molecular biology, ..., physiology, psychology, and social sciences. Beyond these hierarchical structures, an extra “middle world” can be introduced as a bridge between each pair of two adjacent levels. At the lower level, each elementary entity obeys its simple laws; At the middle level, large number of entities interact with each others to form a complex “many-body system” with all dynamical laws become obscure; At the higher level, if the whole complex system is treated as a new entity then it obeys entirely new laws called “emergent phenomena”. The existence of laws for each individual dynamics at the both low and high levels but not the middle one may puzzle readers - Why simple descriptions are legitimate at the both extremes? Are they ruled by independent laws? And what does the role of this middle level play for? To unravel the secrets of this three level architecture being a fundamental logic for complex systems, we should borrow some ideas from physics and physical chemistry.

For complex systems, the three levels correspond nicely to three points of view in physics: *microscopic, mesoscopic*, and *macroscopic*. “Microscopic physics is discrete, governed by probablistic quantum mechanics and by a few simple laws, whereas the laws on macroscopic scale are continuous, deterministic, and manifold” by the French-Armenian physicist R. Balian [8]. Furthermore, as the Israeli physicist Y. Imry stated that [94] “The interest in studying systems in the intermediate size range between microscopic and macroscopic is not only in order to understand the the macroscopic limit and how it is achieved by, say, building up larger and larger clusters to go from the molecule to the bulk. Many novel phenomena exist that are intrinsic to mesoscopic systems." It is now clear that the reasons for the existence of simple laws at microscopic and macroscopic scales are different: The former is due to our *prior knowledge* to describe states of a single elementary entity. The later is an *emergent phenomenon* through limits of *mesoscopic stochastic dynamics*; Taking a limit constitutes a mathematical idealization of the reality. Therefore, the middle world, mesoscopic view, plays crucial roles in bridging those two opposite extremes [86]. Now, we are going to provide an "instruction manual" to build this bridge step by step: (i) introducing variations to the complex systems; (ii) constructing a *sequence* of layers in the middle world.

The first step is about *variations*. Where do they come from? We consider many-body interactions between individuals with large internal degrees of freedom as the origin of variations: If an object has trillions of internal degrees of freedom, no two such objects can be exactly identical in classical dynamic sense; thus variations arise. In statistical mechanics, Gibbs' probabilistic approach to many-body problems gives equilibrium thermodynamics a systematic foundation in terms of his theory of ensemble [80]. The term "ensemble" was coined for the identical copies of the system of interest; the idea of ensemble in physics is corresponding to an independent and identically distributed (i.i.d.) random variables in statistics. Then the probability measure of an event is defined in the original probability space as a prior knowledge and can be verified by the frequency in a large sample size [203]. The existence of distribution laws in different types of ensembles in physics and emergent laws in statistics is illuminating: Introducing the middle world with probabilistic descriptions of variations could be a remedy for us to simplify the complexity. Further, to have the "emergent" laws, we need to be equipped with another concept from mathematical analysis.

The second step is about *asymptotic analysis*, which is a method of describing *emergent behavior* at the higher level through a limiting process. For asymptotic analysis, we need to introduce two fundamental concepts in mathematics: *sequence* and *convergence* in a probabilistic setting. Our three level architecture has been related to the prevailing view of physicists (microscopic-mesoscopic-macroscopic). Not only nature science, in formal science, mathematician has widely applied this idea of level structure to prove limit theorems. For instance, to contruct a *symmetric random walk* $M_k = \sum_{j=1}^{k} X_j$ by

$$X_i = \begin{cases} +1 & p = 1/2, \\ -1 & p = 1/2. \end{cases} \tag{1.1}$$

The Eq. (1.1) gives us a simple rule for each walk at the lower level. (Note that this law has not involved n yet.) Then we construct a *scaled symmetric random walk* $W^{(n)} = (W_t^{(n)})_{t \geq 0}$ by

$$W_t^{(n)} = \frac{1}{\sqrt{n}} M_{nt}, \qquad \text{if } nt \in \mathbb{N}_0, \tag{1.2}$$

where $\sqrt{n}$ is to scale n-steps random walks properly. By constructing the sequence of $W_t^{(n)}$, the "middle world" is divided into n layers. Finally, define a random variable $W_t := \lim_{n\to\infty} W_t^{(n)}$. Then $W_t \sim \mathcal{N}(0, t)$. This emergent random variable, W_t , is known as *Brownian motion*. Through $n \to \infty$, we pass the lower level's characteristics (discrete random walks) on to the higher level (continuous Brownian motion). For the sake of the language of mathematics, we have a clearer picture for the three level structure: (i) A set of equations to represent dynamics for one particle/within one unit of space/in one time step. (ii) A sequence for the equations labeled by n to characterizes the size (scale) of the system (iii) as $n \to \infty$, the convergence gives rise to new laws.

In order to apply the above concept of mathematical limits as an idealization to the reality, here I would like to paraphrase another very insightful statement by P.W. Anderson: [1] "Starting with the microscopic laws, we would have to do two unthinkable things - taking the mathematical limit with infinitely many bodies, and then apply the result to a finite system - before we synthesized the emergent behavior." This statement paints a clear picture of why limit laws are not just abstract concepts but really around us in the real world: Despite nearly all mathematical concepts involve limits, e.g., infinitely many particles, infinitesimal Δx distance, infinitely long time, etc., most of the time when one applies such results to real engineering problems, certain finite number is sufficient, e.g., 1000 is sufficiently many for certain particle system to satisfy thermodynamic

[1] The original text reads [3] "Starting with the fundamental laws and a computer, we would have to do two impossible things - solve a problem with infinitely many bodies, and then apply the result to a finite system - before we synthesized this behavior."

laws, 0.1 centimeter is sufficiently short to quantity an instantaneous change, and 1 minute can be sufficiently long to reach a steady state. The "sufficiency" is corresponding to the "precision" of measurement we can make in the experiment. In reality, there does not exist any perfectly precise measurement. Having inevitably imperfect measurements, we are no longer able to "feel" that tiny difference. Interestingly, emergent laws now become perfect depictions of the world around us.

Based on those two steps (variations and asymptotic analysis) and the illustration of applicability of limit laws to the reality, we can further employ the "instruction manual" to simplify complex systems in different fields. The three level architecture has not only been successfully related to the prevailing view in physics, this concept can also be widely applied to the different fields of science to understand natural science, e.g. biology. In terms of a general scientific method, "separate-levels" turns out to be a relative concept instead of using absolute scales to define "micro-" or "macro-" in physics. For a biological example, a single cell could be either in microscopy or in macroscopy, which totally depends on what system it is relative to. For a biochemist, he/she is interested in a dynamical system of how various concentrations of metabolites change with time (microscopic) in order to describe states of a single cell (macroscopic) [165]; On the other hand, for a cell biologist, he/she is interested in population dynamics of how different types of cells grow and interact with each other (microscopic) to determine the total growth rate and biological functions of a tissue (macroscopic) [93]. In a nutshell, introducing "a sequence of mesoscopic variations" should be the foundation stone for us to understand complex systems, and knowing several novel phenomena from variations intrinsic to the middle world helps us seek a deeper understanding of many intriguing and mysterious concepts in science.

1.1 Emergent limit laws in probability theory

We have mentioned "emergent phenomena" at several places. What is this concept? To make an analogy, the impressionist artists were not trying to paint a realistic picture but rather capture their images without detail. Appreciators change their feelings by looking at the paintings in different distances or from different angles. Those changeable feelings are exactly emergent phenomena under different conditions. It is not hard to find that mathematicians have already portrayed the world in this way for a long time. The *law of large numbers* guarantees an emergent deterministic behavior in the limit of large numbers; the *central limit theorem* tells us about the emergent normal distribution of local fluctuations around the mean on a proper scale; and the *large deviation theory* studies an emergent rate function which depicts the asymptotic probability of rare events approaching to zero. All those three laws are widely applied in science and engineering. On top of that, the *canonical ensemble* to represent the probability of certain quantities in an open system, which has an intrinsic randomness in equilibrium, has also been applied widely in several fields. In Chapter 2, a new limit law of sequence of conditional probability distributions for the canonical ensemble is introduced, which is based on joint work with H. Qian and Y. Zhu [31].

Having the new limit law as a mathematical foundation, a generalizing Gibbisian statistical ensemble theory with its applications to statistical physics is further provided in Chapter 3, which is based on joint work with H. Qian and W. Wang [32]: In the applications of statistical thermodynamics, from condensed matter physics to biophysical chemistry, it is broadly accepted that Gibbs' probabilistic approach to many-body problems is more powerful and efficient than the deterministic Newtonian mechanics [102]. Another parallel canonical ensemble theory by Boltzmann is also discussed with applications to sample frequencies of the phenotype among a population of cells in Chapter 4, which is based on joint work with H. Qian [168]. We now understand why this success: Actually the entire program of Boltzmann and Gibbs are a part of the mathematical theory of limit theorems in probability *beyond* the Law of Large Numbers and the Central Limit Theorems. Indeed, not only macroscopic equilibrium thermodynamics is an emergent phenomenon, the Gibbs' theory itself is a corollary of a limit law from the theory of conditional probability [32].

1.2 Stochastic dynamics of complex systems

In Chapter 5, emergent energetics of stochastic dynamics is established on the basis of kinematics, which is joint work with H. Qian and Y.-J. Yang [170]. Before introducing the concept of *stochastic dynamics* to portray complex systems, it is worth noting another new emerging field - *information theory* - plays an important role in different fields of engineering and science [180] and it has also been increasingly applied in complex systems [36]. Additionally, we need to introduce a more fundamental concept - *a stochastic process* - as the foundation of our theories. A stochastic process is represented by a collection of random variables which is a set of measurable functions indexed by time [95]. Using stochastic differential equations (SDEs) to model a process is based on the relationship between a "random motion" $X(t)$ and its "random velocity" $\dot{X}(t)$. We can see the remarkable resemblance between the role of SDEs in a stochastic dynamical system and the role of *kinematics* in a deterministic dynamical system.

In order to seek a deeper understanding of hidden dynamical laws behind a stochastic process, there are two different approaches: (i) *data-based stochastic models* to unravel the secrets of natural laws (ii) *mechanism-based stochastic models* (stochastic dynamics) to depict complex systems by the language of probability theory. A set of random variables is defined as the *probabilistic representation* for a stochastic process; On top of that, a data-based model for the process is relied on the *statistical representation*; in contrast, mechanism-based stochastic dynamics is characterized by the *informational representation*. The above ternary representation of stochastic change is illustrated in my recent joint work with L. F. Thompson and H. Qian [169]. Now, it is clear that stochastic dynamics is "a stochastic process + mechanism". To make an analogy, in deterministic models, dynamics is "kinematics + natural laws". The following statement provides us an answer for how mechanisms of the stochastic dynamics can be illustrated by the *information* which is centered around the concept of entropy and entropy force in dynamics: Seeking for mechanisms means we want to understand the *causality*. When we talk about the causality, we have already implied the existence of "force" in the system. The force is a well-established quantity in deterministic Newtonian mechanics: $F = ma$ provides us a law of motion in deterministic dynamics. But what is the "force" in stochastic dynamics?

A stochastic dynamics may reach a *stationary state* in the long run and this state can be characterized by an invariant probability distribution. In fact, more often than not this invariant measure is the first "intrinsic, non-trivial measure" of a stochastic dynamical system! This description of stochastic dynamics by the invariant distribution of stationary state is distinct to the stationary states in deterministic dynamics, which conventionally mean fixed points, or limit sets such as a stable limit cycle in the phase space. For the stable limit cycle, an periodic oscillation is a deterministic behavior which is different from the random oscillation as a state of stochastic dynamics in the long run. To connect those two oscillations, stochastic limit-cycle oscillations of a nonlinear system is an example including both of those two types of oscillations, one is periodic from an intrinsic drift and the other one is given by the noises from a diffusion term. In chapter 6, a joint work with Hong Qian for stochastic limit-cycle oscillations of a nonlinear system under random perturbations is provided for further discussions in detail.

Part I

ASYMPTOTIC BEHAVIORS OF CONDITIONAL PROBABILITY DISTRIBUTIONS

Chapter 2

ASYMPTOTIC BEHAVIOR OF A SEQUENCE OF CONDITIONAL PROBABILITY DISTRIBUTIONS AND THE CANONICAL ENSEMBLE

This chapter is based on joint work with Hong Qian and Yizhe Zhu [31].

2.1 Introduction

The canonical ensemble with mechanical energy distribution in an exponential form is the centerpiece of equilibrium statistical mechanics. It represents a *weight* for a microstate of a system in thermal equilibrium with its surrounding *heat bath* at a fixed temperature, where the bath is usually considered much larger in comparison. The theory has wide applications from condensed matter physics to biophysical chemistry [40, 16]. In textbooks, there are currently two heuristic justifications for the exponential factor. One is the original derivation by L. Boltzmann in 1877 based on an ideal gas [181], another is based on the notion of a large heat bath and a small system within, extensively discussed by J. W. Gibbs in his 1902 *magnum opus* [80]. After an extensive discussion of the properties of an *invariant measure* including demonstrating it has to be a function of the mechanical energy, however, Gibbs did not attempt to derive the canonical distribution; rather he simply stated that an exponential form "seems to represent the most simple case conceivable".

Boltzmann's derivation was based on the idea of *most probable frequency* under the constraint of given total energy. In the process he recognized the entropy $S = -N\sum_i f_i \log f_i$ from the multinomial distribution, where N is the number of gas molecules, and i represents a distinct molecule state with kinetic energy e_i. This derivation preceded both the modern theory of large deviations [38, 193] as well as the principles of maximum entropy (MaxEnt) championed by E. T. Jaynes [99, 154]. In terms of the contraction principle in the former, Boltzmann computed the large-deviation rate function for a sample frequency conditioned on a given sample mean of energy instead of obtaining the rate function for the random variable. This approach has now been made rigorous under the heading of the *Gibbs conditioning principle* [188, 38]. MaxEnt, on the other hand, plays a pivotal role in information theory and machine learning [98, 1]. In the 1980s, Boltzmann's logic was also rigorously developed into providing a connection between maximum entropy and conditional probability [217, 196].

Gibbs' theory for the canonical distribution was based on the concept of *heat bath*. In [80], he noted that the distribution with the exponential form had "the property that when the system consists of parts with separate energies, the laws of the distribution in the phase of the separate parts are of the same nature". Having energy E_A for the microstate A of the small system and E_B for the microstate B of the heat bath, Gibbs assumed the phase-space distributions follow (i) additivity: $P(A, B) = P(A + B)$ (ii) independence: $P(A, B) = P(A)P(B)$. Under those two assumptions, the only possible probability distribution for A is exponential: $P(A) \propto e^{\lambda E_A}$. Furthermore, all small systems in contact with the same bath share the same parameter λ, which means they are of the "same nature". By assuming that every small system follows the *conjugate distribution laws* (a family of single parameter exponential priors), A. Ya. Khinchin [108] rigorously proved Gibbs' assertion of the common λ and further showed that it is determined by the given total energy.

The aim of the paper is to find a rigorous origin of the exponential weight itself for the canonical distribution from the standpoint of a heat bath. We were inspired by a very widely used derivation in standard statistical physics textbooks - based on Taylor's expansion of the entropy function of a heat bath [120, 92, 134]. The present work formulates this approach rigorously in probabilistic terms and then gives a proof. We indeed have obtained a rather general new mathematical theorem. The results can be applied back to particular scenarios in statistical physics under corresponding assumptions. Our theorems have clarified the notions of additivity, independency, and the vague "same natures of systems". The last is actually a corollary of the existence and uniqueness of a single parameter in the exponential form of the canonical distribution, and independency is equivalent to additivity of energy functions of two systems during the map from a phase space to its corresponding energy space. We shall emphasize that independency of two systems is a special case of our theorem; the parameter then only depends on fluctuations of the heat bath but independent of the small system.

Our results are obtained based on two mathematical ideas: *conditional probability* and *asymptotics*. We use a *Gedankenexperiment* to illustrate the crucial role of the former - *conditional probability* - in our theorems: Let $Z := X + Y$, where X is a random variable for some function (e.g., energy) of a subsystem and Y describes the same quantity in the heat bath. If one is only interested in the *static statistics* of X, there is a way to set up an experiment: Let $Z(t)$ be a fluctuating total mechanical energy as a function of time, and its distribution has a support on $D \subseteq \mathbb{R}^+$, but one selects only those measurements for $X(t)$ that simultaneously have $Z(t) \in I \subseteq D$. In the language of mathematics, this thought experiment is about the conditional probability of $X(t)$ conditioned on the event $Z(t) \in I$. Why is this thought experiment regarding conditional probability very much in line with the physicist's picture of a canonical ensemble? The answer is in the idea of time-scale separation, which involves three different time scales. The first time scale is for the subsystem $X(t)$ to reach its equilibrium, the second time scale is to restrict the total system $Z(t)$ to be fluctuating inside a finite interval I, and the third time scale is for $Z(t)$ to reach its equilibrium. And the first one is much shorter than the second one, which is much shorter than the third one. Based on this framework of time-scale separation, the canonical ensemble is the statistical ensemble that represents the possible outcomes of the system of interest on the second time scale, i.e., when the subsystem has reached its equilibrium, but the total system is still "constrained" in a certain interval.

In fact, having its own stationary distribution of the total system (if it evolves long enough) is very significant for the theory of conditional probability for two reasons: (1) knowing the fluctuation of the large system is necessary to define the conditional probability mathematically and (2) to perturb the given condition of the total system to see how it has effects on the subsystem is the essence of our theory of the canonical distribution. In other words, even though the original problem is only about the behavior of $X(t)$ when $Z(t) \in I$, if we have more information of $Z(t)$ outside of I, we are able to seek a deeper understanding of the original problem. Not only for the canonical ensemble, this idea of treating a given constraint (parameter) as a variable with distribution has also been widely used in many other fields, for example, comparing the quenched and annealed invariance principles for the random conductance model [10], and in studying the initial-condition naturalness in the case of statistical mechanics [206].

Mathematically using conditional probability to understand Gibbs measure has a long history, see O. E. Lanford [121], O. A. Vasicek [201], H. O. Georgii [78], and H. Touchette [195]. In particular, on the basis of Boltzmann's logic, using asymptotic conditional probability to describe the canonical ensemble has been well-established through the Gibbs conditioning principle [188, 38]. More discussion of this is provided in Section 2.2 for a contradistinction with our own work. In brief, the Gibbs conditioning principle addresses this question: Given a set $A \in \mathbb{R}$ and a constraint $Z_n \in A$, what are the limit points of the conditional probability

$$\mathbb{P}(X_1 \le x \mid Z_n \in A) \quad \text{as } n \to \infty \,? \tag{2.1}$$

In Equation (2.1), $Z_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, where X_i are independent and identically distributed random variables (i.i.d. random variables). We can identify that (2.1) is very similar to our setup for the canonical distribution if we consider $Z_n := \frac{X_1}{n} + Y_n$, where $Y_n = \frac{1}{n}\sum_{i=2}^{n} X_i$ is the measurable function of the heat bath in our approach. However, Y_n in our setup could be defined in a much more general way: we only require that Y_n converges to some random variable Y in distribution (or the law of Y_n satisfies a large deviation principle) rather than has a special form as the sum of independent and identically distributed random variables. In other words, the present work is not a simple refinement of the Gibbs conditioning principle. Here we give a concrete example to which our theorems can be applied but not the Gibbs conditioning principle: Let $\tilde{\zeta}_n = \xi_1 + \eta_n$ and $\eta_n = \sum_{i=2}^{n} \xi_i$, where $\{\xi_i\}_{i=1}^{n}$ are strongly correlated and not identically distributed, and let ζ_n be $\tilde{\zeta}_n$ with appropriate shifting and scaling such that ζ_n has a limiting distribution (or satisfies a large deviation principle). Subject to these conditions, the Gibbs conditioning principle would not be applicable to find the limit points of the conditional probability

$$\mathbb{P}(\xi_1 \le x \mid \zeta_n \in A) \quad \text{as } n \to \infty. \tag{2.2}$$

The present work will show that the canonical distribution in this non-i.i.d. example could still exist as a good approximation (Corollary 2.3.4) or the limiting distribution (Corollary 2.3.10, Corollary

2.3.12) of the conditional probability (2.2). In fact, the setup for our theorems is very general in statistical mechanics: (i) a subsystem in contact with a relatively large heat bath, which is including but not limited to the model of a sum of many independent and identical subsystems; and (ii) the subsystem and the heat bath can have weak or strong interaction.

Back to Equation (2.1), it seems that either using the Gibbs conditioning principle or using our approach to derive the canonical distribution, both sides are asking a very similar question: what is the asymptotic behavior of a conditional probability? However, based on the more general setup of the conditional probability, our approach to the asymptotic behavior of this conditional probability is very different from the Gibbs conditioning principle. For the Gibbs conditioning principle, it transforms the original problem to a sampling problem: what are the limit points of

$$\mathbb{E}\left[L_n \mid L_n \in \Gamma\right] \quad \text{as } n \to \infty\,? \tag{2.3}$$

In Equation (2.3), $L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$ is the corresponding empirical measure for Z_n and $\Gamma = \{\gamma : \int x\gamma(dx) \in A\}$ is the corresponding constraint of Z_n. In fact, even though this approach is called the "Gibbs" conditioning principle, its logic exactly follows Boltzmann's derivation of the canonical ensemble. As a consequence of the Gibbs conditioning principle, it provides a mathematical foundation of why using the maximum entropy principle with certain constraint works to find the canonical distribution [217, 196].

On the other hand, our approach is direct to find the asymptotic behavior of conditional probability (2.1) on the basis of two things: (i) a measurable function of the subsystem is asymptotically small relative to the function of the whole and (ii) the distribution of the measurable function of the heat bath converges to a limiting distribution by appropriate shifting and scaling. Intuitively, under this framework, the distribution of the measurable function of the subsystem shall consist of its unconditional distribution and a weight from a linear approximation of the limiting distribution of the measurable function of the heat bath. As we mentioned above, our approach follows Gibbs' theory for the canonical distribution, which involved the idea of "heat bath" that contributes a "bias" to the system. The common point of our approach and the Gibbs conditional principle is that both sides started with a very similar question of fundamental importance in statistical mechanics and adopted the concept of conditional probability to describe that problem. However, the method of solving the problem on each side has a very different philosophy, the Gibbs conditional principle is about counting statistics by Boltzmann's logic, and ours is inspired by the idea of a heat bath from Gibbs.

Besides the conditional probability, we adopt a very powerful mathematical technique in our theory: asymptotics. Indeed, asymptotics is not only a mathematical technique but also the essence of statistical mechanics. The purpose of statistical mechanics is to derive equilibrium properties of a macroscopic system with enormous numbers of molecules N and occupying a very large volume V, then that macroscopic equilibrium thermodynamics is an *emergent phenomenon* in the limiting case when $N \to \infty$ and $V \to \infty$. Following on from this concept, we shall show that the emergence of an exponential factor in the canonical ensemble is also a result of a limit law according to the probability theory. Take an analogy, our limit theorem is to the exponential form of the canonical distribution what the central limit theorem is to a normal distribution. As with every limit theorem, we have to define how our assumptions depend on n carefully. In our work, as n increases, a measurable function of the subsystem becomes "relatively small" compared with the total system. Based on this main assumption, we obtain two significant results: (i) For a sufficiently large n, a conditional distribution can be well-approximated by its unconditional distribution weighted by an exponential factor, and (ii) a sequence of conditional distributions converges to a limit which is the unconditional distribution weighted by a unique exponential factor.

We obtain two theorems regarding the first result in Section 2.3.2, and they provide the existence of the canonical distribution when a system is contained in a finitely large total system (n is sufficiently large). Furthermore, we obtain two limit theorems regarding the second result in Section 2.3.3, and they provide the existence of a unique canonical distribution when the system is contained in an infinitely large total system ($n \to \infty$). In comparison with Section 2.3.3, Section 2.3.2 only requires weaker conditions, but the exponential form in the canonical distribution may not be unique since there could be more than one sequence having the same asymptotic behavior. On the other hand, Section 2.3.3 requires stronger conditions, but it gives us a unique canonical distribution in the limit, and this distribution can be applied back to approximate the conditional probabilities for all finitely large n. This result can be regarded as an example that the limit theorems from probability predict the laws of nature. Here, we would like to quote from P. W. Anderson [3] "Starting with the fundamental laws and a computer, we would have to do two impossible things - solve a problem with infinitely many bodies, and then apply the result to a finite system - before we synthesized this behavior." Our idea echos Anderson's view: To find the limiting behavior of a sequence of conditional probability distributions and apply it back to the distribution of a subsystem contained in a finitely large total system with some fluctuations, and this is how it is used as a scientific theory.

2.1.1 The equivalence of ensembles

Our work is another way to consider the theory of equivalence of ensembles. As far as we know, Khinchin's derivation of the canonical ensembles in 1949 [108] for a subsystem of a large isolated system by a local central limit theorem was the origin of the equivalence of ensembles. Then Dobrushin and Tirozzi in 1977 [42] extended Khinchin's result from a classical idea gas to a Gibbs random field. In 1979, Martin-Löf [133, 134] further related the microcanonical, canonical, and grand canonical ensembles in the thermodynamic limit when the volume of classical lattice systems tends to infinity. In the 1990s, beyond the scale of the central limit theorem, Deuschel et al. [39] and Georgii [77] showed the equivalence of ensembles on the scale of the large deviation principle. Tasaki [191] recently established the equivalence on the level of local states for large but finite quantum spin systems. A comprehensive introduction to infinite-volume Gibbs measures can be found in Chapter 6 in the textbook by Friedli and Velenik [66], and the discussion of the equivalence of ensembles is in Section 6.14.1.

Recently, a full survey of the equivalence of ensembles at the levels of thermodynamic, macrostates, and measures was presented by Touchette [195]. We shall note that discussions on the equivalence of ensembles at the thermodynamics level can also be traced back to the textbook on statistical mechanics by Hill [87]. In the book, Hill showed the thermodynamic equivalence of ensembles for systems having only a single most probable energy value. In Touchette's recent work, the equivalence was extended to other macrostates, e.g., the mean magnetization of a spin system. This extension was given by the superposition of a mixture of microcanonical ensembles to represent the canonical ensemble of macrostates. Under certain conditions, the equivalence at the macrostate level and the equivalence at the measure level are equivalent. In the language of modern probability, the correspondence between the equivalence at the macrostate level and the equivalence at the measure level is by the Portmanteau theorem [17] for equivalent statements of weak convergence of measures. We shall emphasize that the conditional distribution of the state of a small subsystem converging to the canonical distribution becomes a corollary of the equivalence of ensembles between the microcanonical ensemble and canonical ensemble at the measure level based on the assumption that the state of a small subsystem is chosen at random with a *uniform distribution* in the large whole system.

The essential difference between our approach and the previous approaches for equivalence of ensembles is that we don't assume a uniform distribution of the state of a small subsystem in the large container. This assumption is equivalent to say that the heat bath (the large container - the subsystem) has to be considered as identical copies of the subsystems, which was usually given in the previous work for classical ideal gas systems or Gibbs random fields. In contradistinction to this assumption, our theorems treat the subsystem and its heat bath as two random variables via a measurable function, i.e., we only care about the effect of the "whole" heat bath on the subsystem with respect to that function.

We want to indicate that applying our mathematical theory to physics is new and original since it extends the applicability of the canonical ensemble: by the pushforward measure (up to a prefactor) via a measurable function, we can derive the canonical distribution of a subsystem without assuming a uniform structure of the whole system. For example, we can apply our results to approximate the distribution of certain measurable functions of a small defect within a material. We only require the subsystem (the defect) is small relative to its heat bath with respect to the values of the measurable functions, which is different from treating the heat bath as infinitely large n copies of the subsystem, interacting or not, in order to apply the microcanonical ensemble to the canonical ensemble. In biophysics, our theory can predict the distribution of side-chain conformational variations in protein structure [24, 212, 136]. Proteins in general have a non-uniform structure, so the canonical distribution of side-chain conformational variations can be justified by our theory but not the other approaches based on a uniform structure of the whole system.

We further generalize our theorem to a model when a subsystem and its heat bath have *strong interaction* (the function is not additive), which is beyond the weakly interacting system (the function is approximately additive). This result is new in both physics and mathematics, as a theory for the *Gibbs conditioning principle* for strongly correlated systems. The present work formulates our theory rigorously in probabilistic terms in Section 2.3, and then gives a proof in Section 2.4.

In Section 2.5.1, we apply our theory to concrete examples in statistical mechanics, under two situations when a subsystem and its heat bath are independent or strongly correlated. Since our theory also provides a sharp and precise bound of the convergence rate of conditional probabilities, we use it to approximate the conditional Poisson distribution in Section 2.5.2. To build a connection with the equivalence of ensembles using the techniques of the large deviation principle (LDP) [123] and the central limit theorem (CLT) [42], we applied our theory back to particular scenarios when the heat bath can be treated as a sum of identical random variables. The LDP in Section 2.5.3 or the CLT in Section 2.5.4 gives us a convergence of a sequence of random variables for the heat bath. Nevertheless, we want to emphasize that our theory does not require the LDT or the CTL in general. By proper scaling and shifting, if there exists a convergence of the heat-bath random variable with a smooth limiting distribution, our theory is still applicable. In Section 2.5.5, we provide a precise formulation of what a temperature bath is in probabilistic terms.

2.1.2 Organization of the paper

We provide some useful theorems and definitions and explain our motivation in this problem in Section 2.2. In Section 2.3 we state and explain our main results. Proofs of the main results are provided in Section 2.4. In Section 2.5 we present several applications of our main theorems.

Notations

Throughout the paper, we will adopt the notations $a_n = o(b_n)$ when $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$, and $a_n = O(b_n)$ when $|a_n/b_n|$ is bounded by some constant $C > 0$.

For a set Ω, we use $C(\Omega)$ to represent the set of all continuous real functions on Ω, $C_b(\Omega)$ to represent the set of all bounded continuous functions on Ω, and $C^k(\Omega)$ to represent the set of all functions with continuous derivatives of order k on Ω.

We sometimes use brief notations of probabilities in our proofs, e.g., $P_{X_n|Z_n}(x; I) = P(X_n = x \mid Z_n \in I)$. We always use X_n, Y_n, Z_n to denote sequences of random variables, whose definitions might change in different theorems, but we will give their exact definitions before stating the theorems.

2.2 Preliminaries

2.2.1 Maximum entropy and conditional probability

We first recall the following classical results. Here we don't specify the regularity conditions in the statements of the two theorems below. For more details, see the original references.

Theorem 2.2.1 ([217]). *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed (i.i.d.) random variables with continuous density $f(x)$, then under appropriate regularity conditions, we have*

$$\lim_{n\to\infty} P\left(X_1 \leq x \mid S_n = n\mu + c_n\right) = P\left(X_1 \leq x\right), \tag{2.4}$$

where $S_n := X_1 + X_2 + \cdots + X_n$, $\mu := \mathbb{E}[X_1]$, $s_n^2 := \mathrm{Var}\,[S_n]$, and $c_n = O(s_n)$.

Theorem 2.2.2 ([196]). *Let $\{X_n\}_{n\in\mathbb{N}}$ and S_n follow definitions in Theorem 2.2.1. Let $\alpha \in \mathbb{R}$ and let $f(x)$ be the density function of X_1, then under appropriate regularity conditions,*

$$\lim_{n\to\infty} P\left(X_1 \leq x \mid S_n = n\alpha\right) = \left(\int_{-\infty}^{x} e^{\lambda s} f(s) ds\right) / c(\lambda), \tag{2.5}$$

in which

$$c(\lambda) = \mathbb{E}\left[e^{\lambda X_1}\right] < \infty \quad \textit{and} \quad \alpha = \left(\int x e^{\lambda x} f(x) dx\right) / c(\lambda). \tag{2.6}$$

Note that the parameter λ is determined by the constraint

$$\alpha = \left(\int x e^{\lambda s} f(x) dx \right) / c(\lambda), \tag{2.7}$$

and the density $g(x) = e^{\lambda x} f(x)/c(\lambda)$ maximizes the entropy relative to the density $f(x)$ of X_1 given by

$$H(X_1) = - \int g(x) \log \frac{g(x)}{f(x)} dx, \tag{2.8}$$

with respect to the constraint that

$$\left(\int x g(x) dx \right) = \alpha. \tag{2.9}$$

We see that Theorem 2.2.1 implies the convergence of the conditional probability distribution of X_1 to its unconditional distribution. In this case, the sum of X_i is conditioned on the scale of Gaussian fluctuations: $S_n = n\mu + c_n$, where $n\mu$ is the mean of S_n and c_n is in the order of standard deviation of S_n. On the other hand, we see that Theorem 2.2.2 implies the convergence of the conditional probability distribution of X_1 to the (normalized) product of its unconditional distribution and the maximal entropy distribution $e^{\lambda x}$. The parameter λ is determined by the condition $S_n = n\alpha$, which is on the scale of large deviations when $\alpha \neq \mathbb{E}[X_1]$.

Theorem 2.2.2 is a particular case of the *Gibbs conditioning principle*, which is the meta-theorem [37] regarding the conditional probability of X_i given on the *empirical measure* of an i.i.d. $\{X_i\}_{i=1}^n$

$$L_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \tag{2.10}$$

belongs to some rare event such as

$$\int x L_n(dx) = \frac{1}{n} \sum_{i=1}^{n} X_i = \alpha \quad \text{and} \quad \alpha \neq \mathbb{E}[X_1]. \tag{2.11}$$

Using the empirical measure defined in (2.10) conditioned on the rare event (2.11) to find the limit of conditional probability in Theorem 2.2.2 turns out to be equivalent to find the limit

$$\gamma^* := \lim_{n\to\infty} \mathbb{E}\left[L_n \mid L_n \in \Gamma \right], \quad \Gamma = \{\gamma : \int x \gamma(dx) = \alpha\}. \tag{2.12}$$

By the Gibbs conditioning principle, under appropriate regularity conditions, γ^* minimizes the relative entropy

$$H(\gamma \mid \mu_X) := \int d\gamma \log\left(\frac{d\gamma}{d\mu_X}\right),$$

where $\gamma \in \Gamma$ and μ_X is the law of X_1. In fact, this result implies the limit law derived in Theorem 2.2.2.

One of the most successful approaches to the Gibbs conditioning principle is through the theory of large deviations [188, 37]. This approach involves *Sanov's theorem* [177] that provides the large-deviation rate function of the empirical measure induced by a sequence of i.i.d. random variables and the *contraction principle* [47] that describes how continuous mappings preserve the large deviation principle from one space to another space. In short, these theorems regarding counting and transformation in the theory of large deviations yield the Gibbs conditioning principle and provide the foundation of using the maximum entropy distribution under certain constraints to find the limit of a sequence of conditional probabilities.

2.2.2 *Large deviation theory*

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of i.i.d. absolutely integrable (i.e. $\mathbb{E}|X_1| < \infty$) real random variables with mean $\mu := \mathbb{E}[X_1]$, and let

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2.13}$$

By the *weak law of large numbers*,

$$\overline{X}_n \xrightarrow{P} \mu \quad \text{when } n \to \infty. \tag{2.14}$$

That is, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(|\overline{X}_n - \mu| > \epsilon\right) = 0. \tag{2.15}$$

To study the question how fast this probability tends to zero, Harald Cramér obtained the following theorem in 1938:

Theorem 2.2.3 (Cramér's theorem [35]). *Assume that*

$$A(\lambda) := \log \mathbb{E}[e^{\lambda X_1}] < \infty, \quad \lambda \in \mathbb{R}.$$

Then

$$(i) \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[\overline{X}_n \geq y\right] = -\phi(y) \quad \textit{when } y > \mu,$$
$$(ii) \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[\overline{X}_n \leq y\right] = -\phi(y) \quad \textit{when } y < \mu,$$

where ϕ is defined by

$$\phi(y) := \sup_{\lambda\in\mathbb{R}} [y\lambda - A(\lambda)] \quad \textit{for } x \in \mathbb{R}. \tag{2.16}$$

The function A is called the *logarithmic moment generating function*. In the applications of the large deviation theory to statistical mechanics, A is also called the *free energy function* and the function ϕ is called the rate function of large deviations [193]. We can recognize that $\phi(y)$ is the *Legendre transform* of $A(\lambda)$ (A is a convex function). Therefore, $\phi = A^*$ (the convex conjugate of A) and it leads to the following pair of reciprocal equations

$$\frac{dA(\lambda)}{d\lambda} = y \quad \text{if and only if} \quad \frac{d\phi(y)}{dy} = \lambda. \tag{2.17}$$

Now, we can apply this equivalence (2.17) to Theorem 2.2.2: The parameter λ of the maximum entropy distribution $e^{\lambda s}$ is implicitly solved by (2.7), which gives rise to λ determined by

$$\frac{d\log\int e^{\lambda s} f(s)ds}{d\lambda} = \alpha. \tag{2.18}$$

By the definition of $A(\lambda)$ and (2.17) and (2.18), we have

$$\frac{dA(\lambda)}{d\lambda} = \alpha \quad \text{if and only if} \quad \frac{d\phi(\alpha)}{d\alpha} = \lambda. \tag{2.19}$$

Therefore, this result (2.19) shows that λ not only can be determined implicitly by the free energy function A but also can be founded explicitly by the rate function ϕ.

One of our main theorems (Theorem 2.3.9) can be applied to a particular type of heat bath as the sum of i.i.d. random variables (Theorem 2.5.4), then we directly show that λ is uniquely determined by the first derivative of the rate function ϕ given on the condition α. In this case, we apply the large deviation principle directly to the distribution of the heat bath

$$Y_n = \frac{1}{n}\sum_{i=2}^{n} X_i$$

rather than use the large deviation principle for the empirical measure

$$L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}.$$

In fact, the former (our approach) actually follows Gibbs' logic of the canonical distribution through the heat bath method; The later (Gibbs conditioning principle) follows Boltzmann's logic of the canonical distribution through counting statistics. The reason to call the "Gibbs" conditioning principle was in order to comprehend Gibbs' prediction of the canonical distribution from a mathematical standpoint [188], however, in our opinion, it is closer to the idea of Boltzmann's derivation of the canonical distribution.

From our perspective, choosing the maximum entropy distribution to approximate the conditional probability is a natural consequence of the emergence of $e^{\lambda x} f(x)$ when the finite subsystem is contained in an infinitely large system with a value far from its mean. In other words, (normalized) $e^{\lambda x} f(x)$ is the density of the limit of a sequence of conditional probabilities and it maximizes the relative entropy (2.8) as an inevitable corollary from the setup of the heat bath method. In comparison with the Gibbs conditioning principle, our logic provides a very different point of view of why the maximum entropy principle works to find the limit of conditional probabilities. Even though these two approaches have very different philosophies, in terms of mathematics, they are connected by the reciprocal equations (2.17) through the Legendre transform.

2.2.3 Asymptotic behavior of probabilities

In order to define how "good" of an approximation of conditional probability is, we first need to decide which metric we would use in the space of measures. In what follows, let Ω denote a measurable space with σ-algebra $\mathcal{F}$ and let $\mathbb{P}$, $\mathbb{Q}$ denote two probability measures on $(\Omega, \mathcal{F})$.

Definition 2.2.1 (KL-divergence)**.** *For two probability distributions of a continuous random variable, $\mathbb{P}$ and $\mathbb{Q}$, the KL-divergence is defined by*

$$D_{KL}(\mathbb{P} \parallel \mathbb{Q}) := \int_{-\infty}^{+\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx, \tag{2.20}$$

where p, q are the density functions of $\mathbb{P}, \mathbb{Q}$, respectively. For two probability distributions of a discrete random variable, $\mathbb{P}$ and $\mathbb{Q}$, the Kullback-Leibler divergence between them can be written as

$$D_{KL}(\mathbb{P} \parallel \mathbb{Q}) = \sum_{k \in \Omega} P(k) \log\left(\frac{P(k)}{Q(k)}\right), \tag{2.21}$$

where P, Q are the probability mass functions of $\mathbb{P}, \mathbb{Q}$, respectively and Ω is a countable space. By continuity arguments, the convention is assumed that $0 \log \frac{0}{q} = 0$ for $q \in \mathbb{R}$ and $p \log \frac{p}{0} = \infty$ for $p \in \mathbb{R} \backslash \{0\}$. Therefore, the KL-divergence can take values from zero to infinity.

Definition 2.2.2 (total variation)**.** *The total variation distance between two probability measures $\mathbb{P}, \mathbb{Q}$ on a sigma-algebra $\mathcal{F}$ is defined by*

$$\delta(\mathbb{P}, \mathbb{Q}) := \sup_{A \in \mathcal{F}} |\mathbb{P}(A) - \mathbb{Q}(A)|.$$

It's well known that we have the following relation between KL-divergence and total variation by Pinsker's inequality [153]:

$$\delta(\mathbb{P}, \mathbb{Q}) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(\mathbb{P} \parallel \mathbb{Q})}. \tag{2.22}$$

Definition 2.2.3 (convergence of measures in total variation)**.** *Given the above definition of total variation distance, let $\{\mathbb{P}_n\}_{n \in \mathbb{N}}$ be a sequence of measures on $(\Omega, \mathcal{F})$. The sequence is said to converge to a measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ in total variation distance if*

$$\lim_{n \to \infty} \delta(\mathbb{P}_n, \mathbb{P}) = 0$$

and it is equivalent to

$$\lim_{n \to \infty} \sup_{\|f\|_\infty \leq 1} \left| \int f d\mathbb{P}_n - \int f d\mathbb{P} \right| = 0.$$

Definition 2.2.4 (weak convergence of measures)**.** *Let* $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ *be a sequence of probability measures on* $(\Omega, \mathcal{F})$*. We say that* $\mathbb{P}_n$ *converges* weakly *to a probability measure* $\mathbb{P}$ *on* $(\Omega, \mathcal{F})$ *if*

$$\lim_{n\to\infty} \int f d\mathbb{P}_n = \int f d\mathbb{P},$$

for all $f \in C_b(\Omega)$*.*

From the two definitions above, total variation convergence of measures always implies weak convergence of measures.

Definition 2.2.5 (convergence in distribution)**.** *A sequence* $\{X_n\}_{n\in\mathbb{N}}$ *of random variables is said to* convergence in distribution *to the random variable* X *if*

$$\mu_{X_n} \to \mu_X \quad \textit{weakly},$$

in which μ_{X_n} *is the law of* X_n *and* μ *is the law of* X*.*

Even though the KL-divergence is not a metric, by the inequality (2.22), if the KL-divergence of one sequence of measures from another sequence of measures converges to zero, then the two sequences of measures have to converge to zero in total variation. So they must converge to zero weakly. Following this line of implication, in the present work, we start with defining the KL-divergence between two sequences of measures then understand what conditions guarantee it converges to zero. Once we have that, we will attain both strong convergence and weak convergence of the two sequences of measures to zero under those conditions.

Furthermore we mention two classical theorems (see the reference [182]) regarding the convergence of probability distributions which we will use in our proofs.

Theorem 2.2.4 (Berry-Esseen theorem)**.** *Let* X *have mean zero,* $\mathbb{E}[X^2] = \sigma^2$*, and* $\mathbb{E}|X|^3 < \infty$*. Let* $Z_n = (X_1 + \cdots + X_n)/\sqrt{n}\sigma$*, where* $X_1, \cdots, X_n$ *are i.i.d. copies of* X*. Then we have*

$$|\mathbb{P}(Z_n < z) - \mathbb{P}(G < z)| = O\left(\frac{\mathbb{E}|X|^3}{\sqrt{n}}\right) \tag{2.23}$$

for all $z \in \mathbb{R}$*, where* $G \sim N(0,1)$*.*

Theorem 2.2.5 (Slutsky's theorem). *Let $\{Z_n\}_{n\in\mathbb{N}}, \{W_n\}_{n\in\mathbb{N}}$ be sequences of random variables. If Z_n converges in distribution to a random variable X and W_n converges in probability to a constant c, then*

$$Z_n + W_n \to X + c \quad \textit{in distribution}. \tag{2.24}$$

Corollary 2.2.6. *Let X have mean zero, $\mathbb{E}[X^2] = \sigma^2$, and $\mathbb{E}|X|^3 < \infty$. For some finite $k \in \mathbb{N}$, let $W_n = (X_1 + \cdots + X_k)/\sqrt{n}\sigma$ and $Z_n = (X_{k+1} + \cdots + X_{n+k})/\sqrt{n}\sigma$, where $X_1, \cdots, X_{n+k}$ are i.i.d. copies of X. Let $\tilde{Z}_n = Z_n + W_n$, then we have*

$$\tilde{Z}_n \to G \quad \textit{in distribution}, \quad G \sim N(0,1). \tag{2.25}$$

Furthermore,

$$\left|\mathbb{P}\left(\tilde{Z}_n < z\right) - \mathbb{P}\left(G < z\right)\right| = O\left(\frac{\mathbb{E}|X|^3}{\sqrt{n}}\right) \tag{2.26}$$

for all $z \in \mathbb{R}$.

This corollary follows from Theorem 2.2.4 and Theorem 2.2.5. The proof is provided in Appendix 2.6.2.

2.3 Main results

2.3.1 Setup

In Section 2.1 of introduction, we have already provided our philosophy of adopting the *conditional probability* to derive the canonical ensemble. In this section of the main results, we are going to rigorously show: when a measurable function of the subsystem is "small" relative to the whole system, the "canonical distribution" is a "good" approximation of that conditional distribution. For the sake of simplicity, we will use the terms: "subsystem", "heat bath", and "whole system" to represent a measurable function of those systems, respectively. Within this framework, we first need to define three things rigorously:

1. A relatively small subsystem.

2. Canonical probability distributions.

3. Good approximations.

For the definition of (1): a relatively small subsystem, we consider a sequence of conditional densities

$$f_{X|\tilde{Z}_n}(x; E_n), \quad E_n := \mu_n + I/\beta_n, \tag{2.27}$$

where $\tilde{Z}_n := X + \tilde{Y}_n$, X is a nonnegative continuous random variable and $\tilde{Y}_n$ is a sequence of continuous random variables, I is a finite interval and μ_n, β_n are positive sequences. Note that we here use $\tilde{Y}_n, \tilde{Z}_n$ instead of Y_n, Z_n because we will do transformations for $\tilde{Y}_n, \tilde{Z}_n$ later, so Y_n, Z_n will be used to define transformed $\tilde{Y}_n, \tilde{Z}_n$. The formula of E_n is to represent two kinds of transformations that we can do for the interval I: μ_n is the parameter of shifting and β_n is the parameter of scaling. Through different combinations of μ_n and β_n, the given condition of $\tilde{Z}_n$ will be on certain significant scales. For two examples,

1. Assume $\mu_n := \mathbb{E}[\tilde{Z}_n] = n\mu$, μ is a constant and $\beta_n = 1/\sqrt{n}$, then $\tilde{Z}_n$ is conditioned to be inside the interval $E_n = n\mu + \sqrt{n}I$. The interval E_n is then around $\mathbb{E}[\tilde{Z}_n]$ with a scale of the Gaussian fluctuations in central limit theorem.

2. Assume $\beta_n = 1/n$, then $\tilde{Z}_n$ is conditioned to be inside the interval $E_n = n\mu + nI$. The interval E_n is then around $\mathbb{E}[\tilde{Z}_n]$ with a scale of the large deviations.

In our theorems, we will assume that

$$\mathbb{E}[X^j] < \infty, \text{ for some finite } j, \quad \text{and} \quad \beta_n = o(1). \tag{2.28}$$

Therefore, the definition (2.27) of conditional densities is a sequence of densities for the nonnegative continuous random variable X with $\mathbb{E}[X^j] < \infty$ conditioned on the event $\tilde{Z}_n \in E_n$ with $E_n \to \infty$ ($\beta_n \to 0$). In this way, the positive sequence β_n characterizes that the subsystem is relatively "small" to the given condition of the whole system.

Then we will extend our definition of a "small" subsystem to the case when we have discrete random variables. Consider a sequence of conditional probability functions

$$P\big(K = k \mid \tilde{H}_n \in E_n\big), \quad E_n = \mu_n + I/\beta_n, \tag{2.29}$$

where $\tilde{H}_n := K + \tilde{L}_n$, K is a nonnegative discrete random variables and we assume that

$$\mathbb{E}[K^j] < \infty, \text{ for some finite } j, \quad \text{and} \quad \beta_n = o(1), \tag{2.30}$$

and $\tilde{L}_n$ is a sequence of discrete random variables and $\tilde{H}_n := K + \tilde{L}_n$.

For the definition of (2): canonical probability distributions, we are introducing a general form of the canonical probability distribution as follows: Let I be the interval in the setup (2.27). We consider a sequence of functions $\zeta_n : \mathcal{A} \times \mathbb{R} \to \mathbb{R}$, where $\mathcal{A}$ is the set of all finite intervals on $\mathbb{R}$. For the canonical probability distribution of a nonnegative continuous random variable X, its density can be represented by

$$\frac{f_X(x)e^{-\zeta_n(I;x)x}}{\displaystyle\int_{\mathbb{R}^+} f_X(x)e^{-\zeta_n(I;x)x}dx} \quad \text{and} \quad 0 \le \zeta_n(I;x) < \infty, \text{ for all } x \in \mathbb{R}^+. \tag{2.31}$$

Consider a sequence of functions $\hat{\zeta}_n : \mathcal{A} \times \mathbb{R} \to \mathbb{R}$. For the canonical probability distribution of a nonnegative discrete random variable K, it can be represented by

$$\frac{P(K = k)e^{-\hat{\zeta}_n(I;k)k}}{\sum_{k\in S} P(K = k)e^{-\hat{\zeta}_n(I;k)k}} \quad \text{and} \quad 0 \le \hat{\zeta}_n(I;k) < \infty, \text{ for all } k \in S, \tag{2.32}$$

where S is a set of the support of $P(K = k)$.

For the definition of (3): good approximations, "good" is defined by a sufficiently small distance of two distributions in total variation (2.22). In most of our results, we prove that two sequences of distributions converge to zero in KL-divergence, by Pinsker's inequality, it implies those two sequences converge to zero in total variation, i.e., one sequence is a good approximation of the other one.

2.3.2 Approximation of conditional probabilities

Based on the definitions of (1), (2), and (3) in the setup, we provide two approximation theorems to show the existence of the canonical distributions as good approximations of conditional distributions when the subsystem is sufficiently small relative to the whole systems.

Based on the setup (2.27), let $X_n := \beta_n X$ and take $j = 2$ for the assumption (2.28), i.e.,

$$\mathbb{E}[X^2] < \infty, \quad \text{and} \quad \beta_n = o(1). \tag{2.33}$$

Let $a_n := \beta_n^2 \mathbb{E}[X^2]$, hence we have that

$$\mathbb{E}[X_n^2] = a_n,\ a_n = o(1). \tag{2.34}$$

Let $Y_n := \beta_n \left(\tilde{Y}_n - \mu_n\right)$ and $Z_n := X_n + Y_n$. Note that Y_n, Z_n are the linear transformations of $\tilde{Y}_n, \tilde{Z}_n$, respectively; and recall that $\tilde{Z}_n = X + \tilde{Y}_n$ and the parameters of the transformation, β_n, μ_n, are from $E_n = \mu_n + I/\beta_n$ in the conditional density (2.27). Since we assume I is a finite interval in (2.27), we can define it explicitly as $I = [h, h+\delta],\ h, \delta \in \mathbb{R}$ and $\delta > 0$.

Based on the definitions given above, let $\mathbb{P}_I^{(n)}$ be a sequence of probability measures with density functions

$$\frac{f_X(x) e^{-\beta_n \psi_n(I;\beta_n x) x}}{\displaystyle\int_{\mathbb{R}^+} f_X(x) e^{-\beta_n \psi_n(I;\beta_n x) x} dx}, \quad \psi_n(I; \beta_n x) := \left.\frac{\partial \log P\big(Y_n \in [y, y+\delta] \mid X_n = \beta_n x\big)}{\partial y}\right|_{y=h}. \tag{2.35}$$

And let $\mathbb{Q}_I^{(n)}$ be a sequence of probability measures with density functions $f_{X|\tilde{Z}_n}\big(x; E_n\big)$.

Our first theorem for continuous random variables is as follows:

Theorem 2.3.1. *Assume there exist positive constants C_1, C_2, a positive sequence $b_n = o(1)$, and an open interval D such that the following holds:*

1. *For all $x \in \mathbb{R}^+$, $y \in \mathbb{R}$,*

$$\left|\frac{\partial^2 P\big(Y_n \in [y, y+\delta] \mid X_n = x\big)}{\partial y^2}\right| \leq C_1, \quad \left|\frac{\partial^2 \log P\big(Y_n \in [y, y+\delta] \mid X_n = x\big)}{\partial y^2}\right| \leq C_2. \tag{2.36}$$

2. *For all $x \in \mathbb{R}^+$ and every $[y, y+\delta] \subset D$, there exist positive constants δ_1, C_3 depending on y such that*

$$P\big(Y_n \in [y, y+\delta] \mid X_n = x\big) \geq \delta_1, \quad 0 \leq \frac{\partial \log P\big(Y_n \in [y, y+\delta] \mid X_n = x\big)}{\partial y} \leq C_3, \tag{2.37}$$

$$\big|P\big(Y_n \in [y, y+\delta] \mid X_n = x\big) - P\big(Y_n \in [y, y+\delta]\big)\big| \leq b_n P\big(Y_n \in [y, y+\delta]\big). \tag{2.38}$$

3. *For every $[z, z+\delta] \subset D$, there exists a positive constant δ_2 depending on z such that*

$$P\big(Z_n \in [z, z+\delta]\big) \geq \delta_2. \tag{2.39}$$

Given an interval $I \subset D$, then

$$D_{\text{KL}}\left(\mathbb{P}_I^{(n)} \| \, \mathbb{Q}_I^{(n)}\right) = O(a_n + b_n), \tag{2.40}$$

and $\mathbb{P}_I^{(n)}$ satisfies the definition of the canonical probability distributions in (2.31).

Corollary 2.3.2. *By Pinsker's inequality, Theorem 2.3.1 implies that*

$$\delta\left(\mathbb{P}_I^{(n)}, \mathbb{Q}_I^{(n)}\right) = O\left(\sqrt{a_n + b_n}\right).$$

Corollary 2.3.3. *Interpretations of Theorem 2.3.1 for statistical mechanics: the sequence $a_n = o(1)$ represents that the second moment of the function of the subsystem X scaled by the size of the given condition of the whole system asymptotically goes to zero. And the sequence $b_n = o(1)$ represents that X_n and Y_n are asymptotically independent. By our approximation theorem, using the canonical distribution to approximate the conditional distribution results in a very small error $O(\sqrt{a_n + b_n})$ when n is sufficiently large, i.e.,*

1. *The subsystem is small relative to the whole system.*

2. *The subsystem has weak interaction with its surrounding.*

Note that these conditions (1) *and* (2) *echo the physicist's setup of the canonical ensemble in statistical mechanics.*

For Theorem 2.3.1, we require the condition 2.38 and the sequence b_n in that condition is asymptotic to zero. As Remark 2.3.3, it means that the subsystem and the heat bath are asymptotically independent. In the following corollary, we are going to extend Theorem 2.3.1 to the case when the subsystem X_n and its surrounding (the heat bath) Y_n are not asymptotically independent.

Recall that $I = [h, h+\delta],\ h, \delta \in \mathbb{R}$ and $\delta > 0$ and $\mathbb{Q}_I^{(n)}$ is a sequence of probability measures with density functions $f_{X|\tilde{Z}_n}\big(x; E_n\big)$. Let $\hat{\mathbb{P}}_I^{(n)}$ be a sequence of probability measures with density functions

$$\frac{f_X(x)e^{-\beta_n \phi_n(I) x}}{\displaystyle\int_{\mathbb{R}^+} f_X(x) e^{-\beta_n \phi_n(I) x} dx},$$

where

$$\phi_n(I) := \left.\frac{\partial \log P\big(Y_n \in [y, y+\delta] \mid X_n = 0\big)}{\partial y}\right|_{y=h} - \left.\frac{\partial \log P\big(Y_n \in [y, y+\delta] \mid X_n = 0\big)}{\partial x}\right|_{y=h}. \tag{2.41}$$

Corollary 2.3.4. *Assume there exist positive constants C_1, C_2, and an open interval D such that the following holds:*

1. *For all $x \in \mathbb{R}^+$, $y \in \mathbb{R}$,*

$$\left|\partial^{(2)} P\big(Y_n \in [y, y+\delta] \mid X_n = x\big)\right| \le C_1, \quad \left|\partial^{(2)} \log P\big(Y_n \in [y, y+\delta] \mid X_n = x\big)\right| \le C_2, \tag{2.42}$$

   where $\partial^{(2)}$ denotes all the second order partial derivatives.

2. *For all $x \in \mathbb{R}^+$ and every $[y, y+\delta] \subset D$, there exist positive constants δ_1, C_3 depending on y such that*

$$P\big(Y_n \in [y, y+\delta] \mid X_n = x\big) \ge \delta_1, \quad 0 \le \partial^{(1)} \log P\big(Y_n \in [y, y+\delta] \mid X_n = x\big) \le C_3, \tag{2.43}$$

   where $\partial^{(1)}$ denotes all the first order partial derivatives.

3. *For every $[z, z+\delta] \subset D$, there exists a positive constant δ_2 depending on z such that*

$$P\big(Z_n \in [z, z+\delta]\big) \geq \delta_2. \tag{2.44}$$

Given an interval $I \subset D$, then

$$D_{\mathrm{KL}}\left(\hat{\mathbb{P}}_I^{(n)} \| \, \mathbb{Q}_I^{(n)}\right) = O(a_n), \tag{2.45}$$

and $\hat{\mathbb{P}}_I^{(n)}$ satisfies the definition of the canonical probability distributions in (2.31).

The proof of Corollary 2.3.4 basically follows from the proof of Theorem 2.3.1, and we provide the details of the proof in Appendix 2.6.4.

Corollary 2.3.5. *Here we want to emphasize the difference between Theorem 2.3.1 and Corollary 2.3.4: On the one hand, Corollary 2.3.4 requires a stronger condition that those partial derivatives in the conditions* (2.42) *and* (2.43) *have to be bounded both in the x and y directions; however, Theorem 2.3.1 only requires that the partial derivatives in the conditions* (2.36) *and* (2.37) *are bounded in the y direction. On the other hand, Corollary 2.3.4 does not require the condition* (2.38) *in Theorem 2.3.1, which is to define the asymptotic independence between X_n and Y_n. Based on the difference of those conditions, Theorem 2.3.1 and Corollary 2.3.4 give rise to distinct parameters of the exponential factors. The parameter of the exponential factor* (2.41) *in Corollary 2.3.4 includes one additional term which involves the partial derivative with respect to x.*

Note that the parameter of the exponential factor (2.41) *can be rewritten as*

$$\phi_n(I) = \left.\frac{\partial \log P\big(Y_n \in [y, y+\delta]\big)}{\partial y}\right|_{y=h} + \left.\left(\frac{\partial \log C(x,y)}{\partial y} - \frac{\partial \log C(x,y)}{\partial x}\right)\right|_{(x=0, y=h)}, \tag{2.46}$$

where

$$C(x,y) = \frac{P\big(Y_n \in [y, y+\delta] \mid X_n = x\big)}{P\big(Y_n \in [y, y+\delta]\big)}. \tag{2.47}$$

Corollary 2.3.4 with the parameter represented by (2.46) *has a critical interpretation in statistical mechanics: For a system in contact with a heat bath, if the interaction are not weak (i.e. the correlation in mathematical terms does not approach zero), then the effect of this interaction will appear* *in the parameter of the exponential factor as the function of $C(x, y)$ in* (2.46) *for the canonical distribution. This result is different from the standard example in statistical mechanics: in the limit where the interaction goes to zero, the parameter only includes the effect of the fluctuations of heat bath (the first term on the right side of* (2.46)*) without any effect from the correlation (the second term on the right side of* (2.46)*).*

Now we extend our approximation theorem to discrete random variables based on the setup (2.29). Recall that $\tilde{H}_n = K + \tilde{L}_n$ and $E_n = \mu_n + I/\beta_n$ defined in the conditional probability mass function (2.29). Take $j = 2$ for the assumption (2.30), i.e.,

$$\mathbb{E}[K^2] < \infty, \tag{2.48}$$

and by the definition (2.32), we have a set S such that

$$S := \{k \in \mathbb{R} : P(K = k) > 0\}. \tag{2.49}$$

Let $K_n := \beta_n K$ be a sequence of nonnegative discrete random variables and let $a_n := \beta_n^2 \mathbb{E}[K^2]$. By (2.48) and (2.49), we have that

$$\mathbb{E}[K_n^2] = a_n,\ a_n = o(1), \tag{2.50}$$

and a sequence of sets S_n such that

$$S_n := \{\beta_n k \in \mathbb{R} : P(K_n = \beta_n k) > 0\}. \tag{2.51}$$

By shifting with μ_n and scaling with β_n, we can define a linear transformation of $\tilde{L}_n$, $L_n := \beta_n \left(\tilde{L}_n - \mu_n\right)$, and let $H_n := K_n + L_n$. Furthermore, let Y_n be a sequence of continuous random variables and $Z_n := K_n + Y_n$. Based on the given definitions, our second theorem for discrete random variables is as follows:

Theorem 2.3.6. *Assume the following conditions hold:*

1. *All conditions in Theorem 2.3.1 hold for K_n, Y_n, Z_n on an open interval D.*

2. *There exists a set $D' \subset D$ and a positive sequence $c_n = o(1)$ such that for every interval $I' \subset D'$,*

$$\sup_{\beta_n k \in S_n} \left| P\big(K_n = \beta_n k \mid H_n \in I'\big) - P\big(K_n = \beta_n k \mid Z_n \in I'\big) \right| = O(c_n). \tag{2.52}$$

Given an interval $I \subset D'$, then

$$\sup_{k \in S} \left| P\big(K = k \mid \tilde{H}_n \in E_n\big) - B_n P(K = k) e^{-\beta_n \hat{\psi}_n(I;\beta_n k) k} \right| = O\left(c_n + \sqrt{a_n + b_n}\right), \tag{2.53}$$

where

$$\frac{1}{B_n} := \sum_{k \in S} P(K = k) e^{-\beta_n \psi_n(I;\beta_n k) k} \quad \textit{and} \quad \hat{\psi}_n(I;\beta_n k) := \left. \frac{\partial \log P\big(Y_n \in [y, y+\delta] \mid K_n = \beta_n k\big)}{\partial y} \right|_{y=h},$$

and b_n is defined in Condition (2.38) *of Theorem 2.3.1. Furthermore,*

$$B_n P(K = k) e^{-\beta_n \hat{\psi}_n(I;\beta_n k) k} \tag{2.54}$$

satisfies the definition of the canonical probability distribution in (2.32).

Note that the given assumption (1) in Theorem 2.3.6: all conditions in Theorem 2.3.1 hold for K_n, Y_n, Z_n on an open interval D, in which K_n is corresponding to X_n in Theorem 2.3.1; and all conditions defined for "all $x \in \mathbb{R}^+$" in Theorem 2.3.1 become defined for "all $x \in S_n$" for Theorem 2.3.6. In this way, even K_n is a sequence of discrete random variables, all conditions in Theorem 2.3.1 are well-defined.

Corollary 2.3.7. *In Theorem 2.3.1 and Theorem 2.3.6, X and K are defined as a nonnegative random variable. In the following two points, we extend our approximation theorem to the case when X (or K) is bounded from below (shifting property) and the case when X (or K) is a nonpositive random variable (reflection property):*

1. *(Shifting property) Let X be a continuous random variable bounded from below. By change of variables, let $\hat{X}_n := \beta_n(X - C)$, where C is the finite lower bound, since $\beta_n = o(1)$, we still have*

$$\mathbb{E}[\hat{X}_n^2] = o(1). \tag{2.55}$$

In addition, assume that the conditional probability

$$P\big(Y_n \in [y, y+\delta] \mid \hat{X}_n = x\big)$$

satisfies all of the conditions in Theorem 2.3.1, then we can apply Theorem 2.3.1 to obtain the canonical distribution for X. We call this the shifting property *of the canonical distributions. For the discrete random variable K, its canonical probability distribution has this property as well. This* shifting property *can be interpreted as the extension of the cases restricted to nonnegative quantities (e.g., energy and number of molecules) for the canonical ensemble and the grand canonical ensemble in statistical mechanics: the canonical distribution can be generalized to represent the possible values of a function which is* bounded from below *of the subsystem in thermal equilibrium with the heat bath at a* positive temperature *(In Theorem 2.3.1, we choose the condition I such that $0 \leq \psi_n(I; \beta_n x) < \infty$).*

2. *(Reflection property) Let X be a nonpositive continuous random variable. Assume the condition* (2.37) *in Theorem 2.3.1 becomes*

$$P\big(Y_n \in [y, y+\delta] \mid X_n = x\big) \geq \delta_1, \quad -C_3 < \frac{\partial \log P\big(Y_n \in [y, y+\delta] \mid X_n = x\big)}{\partial y} \leq 0, \tag{2.56}$$

for all $x \in \mathbb{R}^-$. And assume all of the other conditions in Theorem 2.3.1 are satisfied, then Theorem 2.3.1 can be applied to an interval $I = [h, h+\delta] \subset D$ such that $-\infty < \psi_n(I; \beta_n x) \leq 0$, for all $x \in \mathbb{R}^-$. We call this reflection property *of the canonical distributions. For the discrete random variable K, its canonical probability distribution has this property as well. Here is our interpretation of this* reflection property *for statistical mechanics: When a given condition I of the whole system gives rise to a negative parameter ($-\infty < \psi_n(I; \beta_n x) \leq 0$) in the exponential weight of the canonical distribution, our approximation theorem can be applied to the case of a nonpositive function of the subsystem. In combination with this property with the shifting property, the canonical distribution can represent the possible values of a function which is* bounded from above *of the subsystem* *in thermal equilibrium with the heat bath at a* negative temperature *(Here we choose the condition I such that $-\infty < \psi_n(I;\beta_n x) \leq 0$).*

2.3.3 Limit theorems for conditional probabilities

In this section, we provide two limit theorems to show that a sequence of conditional distributions converges to a unique canonical distribution by appropriate shifting and scaling, where the convergence is also in a corresponding scaling of the KL-divergence of this sequence of conditional distributions from its limit distribution. In contrast to the section 2.3.2, here we obtain a unique canonical distribution at the appropriate scale when a system is conditioned on an infinitely large total system ($n \to \infty$). It is different from the section 2.3.2 in which we derive the canonical distribution for each finitely large n directly.

Recall that from the section 2.3.2, for a sufficiently large n, we know that $\mathbb{Q}_I^{(n)}$ with density function

$$f_{X|\tilde{Z}_n}(x;E_n)$$

can be well-approximated by $\mathbb{P}_I^{(n)}$ with density function

$$\frac{f_X(x)e^{-\beta_n\psi_n(I;\beta_n x)x}}{\displaystyle\int_{\mathbb{R}^+} f_X(x)e^{-\beta_n\psi_n(I;\beta_n x)x}dx} \quad \text{and} \quad \psi_n(I;\beta_n x) := \left.\frac{\partial \log P\big(Y_n \in [y,y+\delta] \mid X_n = \beta_n x\big)}{\partial y}\right|_{y=h}. \tag{2.57}$$

Note that the parameter of the exponential function $\psi_n(I;\beta_n x)$ in (2.57) depends on n and x.

Through our limit theorems in this section, we show that the sequence of measures $\mathbb{Q}_I^{(n)}$ can be well-approximated by a unique (sequence of) canonical distribution(s) with density function(s)

$$\frac{f_X(x)e^{-\lambda_n(I)x}}{\displaystyle\int_{\mathbb{R}^+} f_X(x)e^{-\lambda_n(I)x}dx} \tag{2.58}$$

in one of the cases:

1. $\lambda_n(I) = \beta_n\psi(I)$, where $\beta_n = o(1)$, and $\psi : \mathcal{A} \to \mathbb{R}$ is a function such that $\mathcal{A}$ is the set of all finite intervals on $\mathbb{R}$, and $0 < \psi(I) < \infty$.

2. $\lambda_n(I) = \varphi(I)$, where $\varphi : \mathcal{A} \to \mathbb{R}$ is a function satisfying $0 < \varphi(I) < \infty$.

Note that $\psi(I)$ and $\varphi(I)$ are independent of x and n in comparison with $\psi_n(I; x)$ in (2.57). One of the main ideas behind the proof of our limit theorems is as follows: Let $\tilde{\mathbb{P}}_I^{(n)}$ be a sequence of probability measures with density functions (normalized) $f_X(x)e^{-\beta_n \psi(I)x}$, and let $\mathbb{P}_I$ be a probability measure with density function (normalized) $f_X(x)e^{-\varphi(I)x}$. With D_{KL} defined as KL-divergence, Case (1) can be considered as

$$D_{\text{KL}}\left(\tilde{\mathbb{P}}_I^{(n)} \parallel \mathbb{Q}_I^{(n)}\right) \to 0 \quad \text{as } n \to \infty; \tag{2.59}$$

Case (2) can be considered as

$$D_{\text{KL}}\left(\mathbb{P}_I \parallel \mathbb{Q}_I^{(n)}\right) \to 0 \quad \text{as } n \to \infty. \tag{2.60}$$

Note that in Case (1), since $\beta_n = o(1)$, the sequence $\lambda_n(I) \to 0$ for any bounded $\psi(I)$. Therefore, we have to scale the distance D_{KL} by some function of β_n to guarantee the uniqueness of $\psi(I)$. More details are provided in Theorem 2.3.8.

Furthermore, we require stronger conditions than the conditions for (2.57) in order to apply Lemma 2.4.2 and Lemma 2.4.3 to the proof of our limit theorems. Here is the essence of those two lemmas: under appropriate regularity conditions, the sequence $\lambda_n(I)$ in (2.58) is uniquely determined by a linear approximation of the following sequence

$$\log\left(\frac{f_{X|\tilde{Z}_n}(x; E_n)}{f_X(x)}\right). \tag{2.61}$$

Therefore, most of the conditions in our limit theorems are required to guarantee that (2.61) is well-approximated by a linear function and the remainder term converges to zero fast enough.

Recall that $X_n := \beta_n X$, $Y_n := \beta_n\left(\tilde{Y}_n - \mu_n\right)$, and $Z_n := X_n + Y_n$, where β_n, μ_n are positive sequences and $\beta_n = o(1)$, and $E_n = \mu_n + I/\beta_n$, $I = [h, h+\delta],\ h, \delta \in \mathbb{R}$ and $\delta > 0$.

Our first limit theorem for Case (1): $\lambda_n(I) = \beta_n \psi(I)$ is as follows

Theorem 2.3.8. *Consider a function $\psi : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$ such that $0 < \psi(I) < \infty$ for the given interval I. Let $\tilde{\mathbb{P}}_I^{(n)}$ be a sequence of probability measures with density functions*

$$\frac{f_X(x)e^{-\beta_n \psi(I)x}}{\displaystyle\int_{\mathbb{R}^+} f_X(x)e^{-\beta_n \psi(I)x}dx}. \tag{2.62}$$

Assume the following conditions hold:

1. *X is a nonconstant random variable with $\mathbb{E}[X^3] < \infty$ and $\dfrac{f_{X|\tilde{Z}_n}(x;E_n)}{f_X(x)}$ is uniformly bounded on $\mathbb{R}^+$.*

2. *$Y_n \to Y$ in distribution. The distribution function of Y is bounded on $\mathbb{R}^+$ and satisfies*

$$\log P\left(Y \in [y, y+\delta]\right) \in C^2(D) \quad \text{and} \quad 0 < \left.\frac{\partial \log P(Y \in [y, y+\delta])}{\partial y}\right|_h < \infty, \tag{2.63}$$

   where D is an open interval containing h.

3. *There exists a sequence of functions $g_n : \mathbb{R} \to \mathbb{R}$ with $\left|g_n(x)e^{-\beta_n \xi x}\right|$ uniformly bounded on $\mathbb{R}^+$ for any $\xi > 0$ and $\mathbb{E}\left[g_n(X)^2\right] \to 0$ such that on $I_n = [0, d_n]$ with $d_n = O\left(\frac{1}{\beta_n}\right)$,*

$$\log\left(\frac{P\left(Y_n \in I - \beta_n x \mid X_n = \beta_n x\right)}{P\left(Z_n \in I\right)}\right) = \log\left(\frac{P\left(Y \in I - \beta_n x\right)}{P\left(Y \in I\right)}\right) + \beta_n g_n(x). \tag{2.64}$$

Then

$$\lim_{n\to\infty} \frac{D_{\mathrm{KL}}\left(\tilde{\mathbb{P}}_I^{(n)} \parallel \mathbb{Q}_I^{(n)}\right)}{\beta_n^2} = 0 \quad \text{if and only if} \quad \psi(I) = \left.\frac{\partial \log P(Y \in [y, y+\delta])}{\partial y}\right|_h.$$

And $\tilde{\mathbb{P}}_I^{(n)}$ satisfies the definition of the canonical probability distributions in (2.31).

Our second limit theorem for Case (2): $\lambda_n(I) = \varphi(I)$ is as follows

Theorem 2.3.9. *Let $\varphi : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$ be a function such that $0 < \varphi(I) < \infty$ for the given interval I. Let $\mathbb{P}_I$ be a probability measure with density function*

$$\frac{f_X(x)e^{-\varphi(I)x}}{\displaystyle\int_{\mathbb{R}^+} f_X(x)e^{-\varphi(I)x}dx}.$$

Assume the following conditions hold:

1. *X is a nonconstant random variable with $\mathbb{E}[X] < \infty$ and $\dfrac{f_{X|\tilde{Z}_n}(x; E_n)}{f_X(x)}$ is uniformly bounded on $\mathbb{R}^+$.*

2. *$Y_n \to \mu$ in probability, for some constant $\mu \notin I$. The sequence of laws of Y_n satisfies a large deviation principle with speed $1/\beta_n$ and rate function $\phi \in C^2(D)$, where D is an open interval containing I, and $-\infty < \phi'(y) < 0$ for all $y \in I$.*

3. *There exists a sequence of functions $r_n : \mathbb{R} \to \mathbb{R}$ with $\left|r_n(x)e^{-\xi x}\right|$ uniformly bounded on $\mathbb{R}^+$ for any $\xi > 0$ and $\mathbb{E}\left[r_n(X)^2\right] \to 0$ such that on $I_n = [0, d_n]$ with $d_n = O\left(\frac{1}{\beta_n}\right)$,*

$$\log\left(\frac{P\left(Y_n \in I - \beta_n x \mid X_n = \beta_n x\right)}{P\left(Z_n \in I\right)}\right) = \log\left(\frac{\exp\left[-\frac{1}{\beta_n}\phi\left(y^* - \beta_n x\right)\right]}{\exp\left[-\frac{1}{\beta_n}\phi\left(y^*\right)\right]}\right) + r_n(x), \tag{2.65}$$

$$y^* = \{y : \inf_{y \in I} \phi(y)\}. \tag{2.66}$$

Then

$$\lim_{n\to\infty} D_{\mathrm{KL}}\left(\mathbb{P}_I \parallel \mathbb{Q}_I^{(n)}\right) = 0 \quad \textit{if and only if} \quad \varphi(I) = -\phi'(y^*).$$

And $\mathbb{P}_I$ satisfies the definition of the canonical probability distributions in (2.31).

The following is the discussion about the circumstances when the condition (2.64) (or the condition (2.65)) for Theorem 2.3.8 (or Theorem 2.3.9) could hold. Here we only discuss the condition (2.64) but it is applied to the condition (2.65) as well. We can consider three circumstances

1. When $Y_n \to Y$ in distribution.

2. When $X_n \to 0$ in probability.

3. When X_n and Y_n are asymptotically independent.

Even thought the condition (2.64) and the combination of circumstances (1) - (3) are not the exact same, they are very close; therefore, these three circumstances provide us an insight regarding three elements of the condition (2.64): Circumstance (1) means the heat bath has a limiting distribution; Circumstance (2) means the subsystem is relatively small in comparison with the whole system; Circumstance (3) means those two systems have weak interaction.

As Corollary 2.3.4 for the approximation theorem 2.3.1, we are going to extend our limit theorems to the case when X_n and Y_n are not asymptotically independent.

Corollary 2.3.10. *Let $G : \mathcal{A} \times \mathbb{R} \to \mathbb{R}$ be a function such that $\mathcal{A}$ is the set of all finite intervals. For the given interval I, $G(I; 0) = 1$ and $\log G(I, \xi) \in C^2(\mathbb{R}^+)$ with respect to ξ. Assume the condition* (2.64) *in Theorem 2.3.8 becomes*

$$\log\left(\frac{P\left(Y_n \in I - \beta_n x \mid X_n = \beta_n x\right)}{P\left(Z_n \in I\right)}\right) = \log\left(\frac{P\left(Y \in I - \beta_n x\right) \cdot G(I; \beta_n x)}{P\left(Y \in I\right)}\right) + \beta_n g_n(x), \tag{2.67}$$

and the other conditions in Theorem 2.3.8 hold. Then

$$\lim_{n\to\infty} \frac{D_{\mathrm{KL}}\left(\tilde{\mathbb{P}}_I^{(n)} \parallel \mathbb{Q}_I^{(n)}\right)}{\beta_n^2} = 0 \quad \textit{if and only if} \quad \psi(I) = \left.\frac{\partial \log P(Y \in [y, y+\delta])}{\partial y}\right|_h - \left.\frac{\partial \log G(I; \xi)}{\partial \xi}\right|_0.$$

And $\tilde{\mathbb{P}}_I^{(n)}$ satisfies the definition of the canonical probability distributions in (2.31).

Corollary 2.3.11. *The function G in* (2.67) *can be considered as an approximation:*

$$\frac{P\left(Y_n \in I - \xi \mid X_n = \xi\right)}{P\left(Y_n \in I - \xi\right)} \approx G(I; \xi),$$

in which the left side is equivalent to the joint probability of X_n, Y_n divided by the products of their marginal probabilities. Therefore, G could represent an estimation of the correlation of X_n and Y_n; in information theory, the function G is closely related to the mutual information between X_n and Y_n.

Corollary 2.3.12. *Let $R : \mathcal{A} \times \mathbb{R} \to \mathbb{R}$ be a function, for the given interval I, $R(I; 0) = 1$ and $\log R(I, \xi) \in C^2(\mathbb{R}^+)$ with respect to ξ. Assume the condition* (2.65) *in Theorem 2.3.9 becomes*

$$\log\left(\frac{P\left(Y_n \in I - \beta_n x \mid X_n = \beta_n x\right)}{P\left(Z_n \in I\right)}\right) = \log\left(\frac{\exp\left[-\frac{1}{\beta_n}\phi\left(y^* - \beta_n x\right)\right] \cdot \left(R(I; \beta_n x)\right)^{\frac{1}{\beta_n}}}{\exp\left[-\frac{1}{\beta_n}\phi\left(y^*\right)\right]}\right) + r_n(x), \tag{2.68}$$

$$y^* = \{y : \inf_{y\in I} \phi(y)\}, \tag{2.69}$$

and the other conditions in Theorem 2.3.9 hold. Then

$$\lim_{n\to\infty} D_{\mathrm{KL}}\left(\mathbb{P}_I \parallel \mathbb{Q}_I^{(n)}\right) = 0 \quad \textit{if and only if} \quad \varphi(I) = -\phi'(y^*) - \left.\frac{\partial \log R(I;\xi)}{\partial \xi}\right|_0 .$$

And $\mathbb{P}_I$ satisfies the definition of the canonical probability distributions in (2.31).

Corollary 2.3.13. *The function R in* (2.68) *can be considered as an approximation:*

$$\frac{P\left(Y_n \in I - \xi \mid X_n = \xi\right)}{P\left(Y_n \in I - \xi\right)} \approx \left(R(I,\xi)\right)^{\frac{1}{\beta_n}} .$$

In comparison with Corollary 2.3.10, when the sequence of laws of Y_n satisfies a large deviation principle with speed $1/\beta_n$, the correlation of the subsystem and its heat bath has to be in $O(R^{\frac{1}{\beta_n}})$ to contribute an additional term in the parameter of the exponential weight. Otherwise, if the correlation is just in $O(R)$ as the order in Corollary 2.3.10, then it has no influence on the canonical distribution.

The proof of Corollary 2.3.10 (Corollary 2.3.12) basically follows from the proof of Theorem 2.3.8 (Theorem 2.3.9). We provide the details of proof in Appendix 2.6.4.

As our approximation theorems in Section 2.3.2, we can extend our limit theorems to discrete random variables, random variables bounded below, and random variables bounded above as follows:

1. Discrete random variables: Theorem 2.3.8 and Theorem 2.3.9 can also be applied to the case when we have a nonnegative discrete random variable K, a sequence of discrete random variables $\tilde{L}_n$, and $\tilde{H}_n := K + \tilde{L}_n$. It is said that the sequence of conditional probabilities $P(K = k \mid \tilde{H}_n \in E_n)$ has a limit (by appropriate scaling)

$$\frac{P(K=k)e^{-\lambda_n(I)k}}{\sum_{k\in S} P(K=k)e^{-\lambda_n(I)k}}. \tag{2.70}$$

   The case of $\lambda_n(I) = \beta_n\psi(I)$ follows from Theorem 2.3.8; The case of $\lambda_n(I) = \varphi(I)$ follows from Theorem 2.3.9. Furthermore, the probability function (2.70) satisfies the definition of the canonical probability distribution in (2.32).

2. Random variables bounded below: As Remark 2.3.7, we can extend those limit theorems to the case when X is bounded below. By change of variable, let $\hat{X}_n := \beta_n(X - C)$, where C is the finite lower bound, we still have

$$\mathbb{E}[(X - C)^j] < \infty,\ j = 1 \text{ or } 3. \tag{2.71}$$

Note that $j = 3$ is for Theorem 2.3.8 and $j = 1$ is for Theorem 2.3.9. In addition, assume

$$\log\left(\frac{P\left(Y_n \in I - \beta_n x \mid X_n = \beta_n x\right)}{P\left(Z_n \in I\right)}\right)$$

satisfies the condition of linear approximation in (2.64) and (2.65), for Theorem 2.3.8 and Theorem 2.3.9, respectively. Then we can apply those limit theorems to obtain a unique canonical distribution of X. Therefore, as the point (1) in Remark 2.3.7, a unique canonical distribution derived by the limit of a sequence of conditional distributions has the "shifting property". For the discrete random variable K, its unique canonical distribution has this property as well.

3. Random variables bounded above: Let X be a nonpositive continuous random variable and the corresponding canonical distribution be a sequence of distributions with density functions

$$\frac{f_X(x)e^{-\lambda_n(I)x}}{\displaystyle\int_{\mathbb{R}^-} f_X(x)e^{-\lambda_n(I)x}dx}, \quad -\infty < \lambda_n(I) < 0. \tag{2.72}$$

When $\lambda_n(I) = \beta_n\psi(I)$, Theorem 2.3.8 can be applied to an interval I such that $-\infty < \psi(I) < 0$; When $\lambda_n(I) = \varphi(I)$, Theorem 2.3.9 can be applied to an interval I such that $-\infty < \varphi(I) < 0$. Therefore, as the point (2) in Remark 2.3.7, a unique canonical distribution derived by the limit of a sequence of conditional distributions has the "reflection property". For the discrete random variable K, its unique canonical distribution has this property as well. This reflection property provides us an explanation of the possibility of *negative temperature*: For some given condition of the whole system which arises a *negative* parameter $(-\infty < \lambda_n(I) < 0)$ in the exponential weight, a unique canonical distribution for a function *bounded from above* of the subsystem emerges as the limit of a sequence of conditional distributions.

2.4 Proofs of main results

2.4.1 Proofs of Theorem 2.3.1 and Theorem 2.3.6

Proof of Theorem 2.3.1

Proof. We first prove for the case: $\{x : f_{X_n}(x) > 0\} = \mathbb{R}^+$. In this case, $P(Z_n \in I \mid X_n = x)$ is well-defined for all $x \in \mathbb{R}^+$. Let

$$I = [h, h+\delta] \subseteq D, \quad I - x := \{y - x : \ y \in I\},$$

with Condition (2.39): for $I \subseteq D$, $P\big(Z_n \in [h, h+\delta]\big) \geq \delta_2$, we can derive the following conditional density by Bayes' theorem

$$f_{X_n|Z_n}(x; I) = \frac{f_{X_n}(x)P(Z_n \in I \mid X_n = x)}{P(Z_n \in I)} = \frac{f_{X_n}(x)P(Y_n \in I - x \mid X_n = x)}{P(Z_n \in I)}, \quad \text{for } x \in \mathbb{R}^+. \tag{2.73}$$

Note that $P(Y_n \in I - x \mid X_n = x) = P(Y_n \in [h - x, h + \delta - x] \mid X_n = x)$. Define

$$G_\delta(y, x) := P\big(Y_n \in [y, y + \delta] \mid X_n = x\big).$$

By Taylor expansion and Condition (2.36), we can expand $G_\delta(h - x, x)$ at (h, x) to get

$$G_\delta(h - x, x) = G_\delta(h, x) - \frac{\partial G_\delta(h, x)}{\partial y} x + \frac{\partial^2 G_\delta(h - \alpha_n x, x)}{2\partial y^2} x^2, \quad \text{for some } \alpha_n \in (0, 1).$$

It implies that

$$\begin{aligned}
&P\big(Y_n \in [h - x, h + \delta - x] \mid X_n = x\big) \\
=&P\big(Y_n \in [h, h+\delta] \mid X_n = x\big) - \left.\frac{\partial P\big(Y_n \in [y, y+\delta] \mid X_n = x\big)}{\partial y}\right|_{y=h} \cdot x + r_n(x)x^2 \\
=&P\big(Y_n \in [h, h+\delta] \mid X_n = x\big) \left[1 - \psi_n(I; x) \cdot x + \frac{r_n(x)x^2}{P\big(Y_n \in [h, h+\delta] \mid X_n = x\big)}\right] \\
=&P\big(Y_n \in [h, h+\delta] \mid X_n = x\big) \left[e^{-\psi_n(I;x)x} - \frac{(\psi_n(I; x)x)^2 e^{-\gamma_n \cdot \psi_n(I;x)x}}{2} + \frac{r_n(x)x^2}{P\big(Y_n \in [h, h+\delta] \mid X_n = x\big)}\right] \\
=&P\big(Y_n \in [h, h+\delta] \mid X_n = x\big) \left[e^{-\psi_n(I;x)x} + k_n(x)x^2\right],
\end{aligned} \tag{2.74}$$

where

$$\psi_n(I;x) = \frac{\partial \log P\big(Y_n \in [y, y+\delta] \mid X_n = x\big)}{\partial y}\bigg|_{y=h}, \tag{2.75}$$

$$r_n(x) = \frac{1}{2}\frac{\partial^2 P\big(Y_n \in [y, y+\delta] \mid X_n = x\big)}{\partial y^2}\bigg|_{y=h-\alpha_n x}, \tag{2.76}$$

$$k_n(x) = \frac{r_n(x)}{P\big(Y_n \in [h, h+\delta] \mid X_n = x\big)} - \frac{\psi_n(I;x)^2 e^{-\gamma_n \cdot \psi_n(I;x)x}}{2}, \tag{2.77}$$

and we have applied Taylor's expansion

$$e^{y_n} = 1 + y_n + \frac{(y_n)^2 e^{\gamma_n y_n}}{2}, \quad \text{for some } \gamma_n \in (0,1) \quad \text{and} \quad y_n := \psi_n(I;x)x$$

to the third equation in (2.74). Note that by Condition (2.37),

$$0 \le \psi_n(I;x) \le C_3, \tag{2.78}$$

and by Conditions (2.36) and (2.37), for all $x \in \mathbb{R}^+$, $k_n(x)$ is uniformly bounded. Therefore, by the results of (2.73) and (2.74), for all $x \in \mathbb{R}^+$, we obtain that

$$f_{X_n|Z_n}\big(x;I\big) = \frac{f_{X_n}(x)P(Y_n \in I \mid X_n = x)(e^{-\psi_n(I;x)x} + k_n(x)x^2)}{P(Z_n \in I)}. \tag{2.79}$$

In the following, we will use brief notations

$$P_{Y_n|X_n}\big(I;x\big) := P(Y_n \in I \mid X_n = x), \quad P_{Z_n}\big(I\big) := P(Z_n \in I).$$

First, we let

$$A_n := \frac{1}{\displaystyle\int_{\mathbb{R}^+} f_{X_n}(x)e^{-\psi_n(I;x)x}dx}. \tag{2.80}$$

Since $\int_{\mathbb{R}^+} f_{X_n}(x)dx = 1$, from (2.78), we have $\displaystyle\int_{\mathbb{R}^+} f_{X_n}(x)e^{-\psi_n(I;x)x}dx \le 1$, hence $A_n \ge 1$ for all $n \ge 1$. By definition $X_n = \beta_n X$, $\beta_n \to 0$, we also have

$$\lim_{n\to\infty} \frac{1}{A_n} = \lim_{n\to\infty} \int_{\mathbb{R}^+} f_{X_n}(x)e^{-\psi_n(I;x)x}dx = \lim_{n\to\infty} \int_{\mathbb{R}^+} f_X(x)e^{-\psi_n(I;\beta_n x)\beta_n x}dx = 1$$

by the dominated convergence theorem, so A_n is uniformly bounded from above and from below.

Recall the definition of KL-divergence from (2.20), and the definitions of $\mathbb{P}_I^{(n)}$ and $\mathbb{Q}_I^{(n)}$ from (2.35), we have

$$
\begin{aligned}
D_{\mathrm{KL}}\left(\mathbb{P}_I^{(n)} \parallel \mathbb{Q}_I^{(n)}\right) &= \int_{\mathbb{R}^+} A_n f_X(x) e^{-\psi_n(I;\beta_n x)\beta_n x} \log\left(\frac{A_n f_X(x) e^{-\psi_n(I;\beta_n x)\beta_n x}}{f_{X|\tilde{Z}_n}(x, E_n)}\right) dx \\
&= \int_{\mathbb{R}^+} A_n f_{X_n}(x) e^{-\psi_n(I;x)x} \log\left(\frac{A_n f_{X_n}(x) e^{-\psi_n(I;x)x}}{f_{X_n|Z_n}(x, I)}\right) dx \qquad (2.81) \\
&= \left| \int_{\mathbb{R}^+} A_n f_{X_n}(x) e^{-\psi_n(I;x)x} \log\left(\frac{f_{X_n|Z_n}(x, I)}{A_n f_{X_n}(x) e^{-\psi_n(I;x)x}}\right) dx \right|. \qquad (2.82)
\end{aligned}
$$

(2.81) is obtained by the change of variables $X_n = \beta_n X$ and the scale invariant property of the KL-divergence. (2.82) is true because the KL-divergence is nonnegative. With (2.73), the right hand side in (2.82) can be written as

$$
\begin{aligned}
&\left| \int_{\mathbb{R}^+} A_n f_{X_n}(x) e^{-\psi_n(I;x)} \log\left(\frac{f_{X_n}(x) P_{Y_n|X_n}(I-x;x)}{P_{Z_n}(I)} \cdot \frac{1}{A_n f_{X_n}(x) e^{-\psi_n(I;x)x}}\right) dx \right| \\
=& \left| \int_{\mathbb{R}^+} A_n f_{X_n}(x) e^{-\psi_n(I;x)} \log\left(\frac{P_{Y_n|X_n}(I-x;x)}{P_{Y_n|X_n}(I;x) e^{-\psi_n(I;x)x}} \cdot \frac{P_{Y_n|X_n}(I;x)}{P_{Z_n}(I) A_n}\right) dx \right| \\
\leq& \left| \int_{\mathbb{R}^+} A_n f_{X_n}(x) e^{-\psi_n(I;x)} \log\left(\frac{P_{Y_n|X_n}(I;x)}{P_{Z_n}(I)\, A_n}\right) dx \right| \\
&+ \left| \int_{\mathbb{R}^+} A_n f_{X_n}(x) e^{-\psi_n(I;x)} \log\left(\frac{P_{Y_n|X_n}(I-x;x)}{P_{Y_n|X_n}(I;x) e^{-\psi_n(I;x)x}}\right) dx \right|. \qquad (2.83)
\end{aligned}
$$

From the expression of $f_{X_n|Z_n}(x;I)$ in (2.79), we have the following identity

$$
\begin{aligned}
1 &= \int_{\mathbb{R}^+} f_{X_n|Z_n}(x;I) dx \\
&= \frac{\int_{\mathbb{R}^+} f_{X_n}(x) e^{-\psi_n(I;x)x} P_{Y_n|X_n}(I;x) dx}{P_{Z_n}(I)} + \frac{\int_{\mathbb{R}^+} f_{X_n}(x) k_n(x) x^2 P_{Y_n|X_n}(I;x) dx}{P_{Z_n}(I)}. \qquad (2.84)
\end{aligned}
$$

For the second term in (2.84), Conditions (2.36) and (2.37) imply that $P_{Y_n|X_n}(I;x)$ and $k_n(x)$ are uniformly bounded and Condition (2.39) implies that $P_{Z_n}(I)$ is uniformly bounded from below.

Then by the assumption $\mathbb{E}[X_n^2] = a_n$, the first term in (2.84) satisfies

$$\left|\frac{\int_{\mathbb{R}^+} f_{X_n}(x)e^{-\psi_n(I;x)x}P_{Y_n|X_n}\big(I;x\big)dx}{P_{Z_n}\big(I\big)}\right| = \left|\frac{\int_{\mathbb{R}^+} f_{X_n}(x)k_n(x)x^2P_{Y_n|X_n}\big(I;x\big)dx}{P_{Z_n}\big(I\big)} - 1\right| = 1 + O(a_n). \tag{2.85}$$

With Condition (2.37), (2.85) implies

$$\frac{1}{P_{Z_n}\big(I\big)A_n} = \frac{\int_{\mathbb{R}^+} f_{X_n}(x)e^{-\psi_n(I;x)x}dx}{P_{Z_n}\big(I\big)} \leq \frac{1}{\delta_1} + O(a_n). \tag{2.86}$$

By Conditions (2.37) and (2.38): $\left|P_{Y_n|X_n}\big(I;x\big) - P_{Y_n}\big(I\big)\right| \leq b_nP_{Y_n}\big(I\big)$ with $b_n \to 0$, therefore

$$P_{Y_n}\big(I\big) \leq \frac{P_{Y_n|X_n}\big(I;x\big)}{1-b_n} \leq K_1 \tag{2.87}$$

for some constant $K_1 > 0$. With (2.86) and (2.87) and recalling the definition of A_n in (2.80), we have

$$\begin{aligned}
&\left|\frac{\int_{\mathbb{R}^+} f_{X_n}(x)e^{-\psi_n(I;x)x}P_{Y_n|X_n}\big(I;x\big)dx}{P_{Z_n}\big(I\big)} - \frac{P_{Y_n}\big(I\big)}{P_{Z_n}\big(I\big)A_n}\right| \\
=&\left|\frac{\int_{\mathbb{R}^+} f_{X_n}(x)e^{-\psi_n(I;x)x}\left(P_{Y_n|X_n}\big(I;x\big) - P_{Y_n}\big(I\big)\right)dx}{P_{Z_n}\big(I\big)}\right| \\
\leq&\frac{\int_{\mathbb{R}^+} f_{X_n}(x)e^{-\psi_n(I;x)x}\left|P_{Y_n|X_n}\big(I;x\big) - P_{Y_n}\big(I\big)\right|dx}{P_{Z_n}\big(I\big)} \\
\leq&\frac{b_nP_{Y_n}\big(I\big)\int_{\mathbb{R}^+} f_{X_n}(x)e^{-\psi_n(I;x)x}dx}{P_{Z_n}\big(I\big)} = \frac{b_nP_{Y_n}\big(I\big)}{P_{Z_n}\big(I\big)A_n} \leq K_1b_n\left(\frac{1}{\delta_1} + O(a_n)\right) = O(b_n).
\end{aligned} \tag{2.88}$$

And similarly,

$$\left|\frac{P_{Y_n|X_n}\big(I;x\big)}{P_{Z_n}\big(I\big)A_n} - \frac{P_{Y_n}\big(I\big)}{P_{Z_n}\big(I\big)A_n}\right| \leq \frac{b_nP_{Y_n}\big(I\big)}{P_{Z_n}\big(I\big)A_n} = O(b_n). \tag{2.89}$$

By the triangle inequality, from (2.85), (2.88) and (2.89), we have

$$\frac{P_{Y_n|X_n}(I;x)}{P_{Z_n}(I)A_n} = 1 + O(a_n + b_n).$$

Since $\log(1+x) \le x$ for all $x > -1$, for sufficiently large n, we have

$$\log\left(\frac{P_{Y_n|X_n}(I;x)}{P_{Z_n}(I)A_n}\right) = O(a_n + b_n). \tag{2.90}$$

Note that the term $O(a_n + b_n)$ in (2.90) is independent of x. Therefore, for the first term in (2.83) we have

$$\begin{aligned}
&\left|\int_{\mathbb{R}^+} A_n f_{X_n}(x) e^{-\psi_n(I;x)x} \log\left(\frac{P_{Y_n|X_n}(I;x)}{P_{Z_n}(I)A_n}\right) dx\right| \\
\le& \sup_x \left|\log\left(\frac{P_{Y_n|X_n}(I;x)}{P_{Z_n}(I)A_n}\right)\right| \int_{\mathbb{R}^+} A_n f_{X_n}(x) e^{-\psi_n(I;x)x} dx \\
=& \sup_x \left|\log\left(\frac{P_{Y_n|X_n}(I;x)}{P_{Z_n}(I)A_n}\right)\right| \cdot 1 = O(a_n + b_n).
\end{aligned} \tag{2.91}$$

Define

$$\hat{G}_\delta(y,x) := \log P\big(Y_n \in [y, y+\delta] \mid X_n = x\big).$$

Then by Taylor expansion and the conditions (2.36), (2.37), we can expand $\hat{G}_\delta(h-x,x)$ at (h,x) to get

$$\begin{aligned}
\hat{G}_\delta(h-x,x) &= \hat{G}_\delta(h,x) - \frac{\partial \hat{G}_\delta(h,x)}{\partial y}x + \frac{\partial^2 \hat{G}_\delta(h-\hat{\alpha}_n x, x)}{2\partial y^2}x^2, \quad \text{for some } \hat{\alpha}_n \in (0,1), \\
&= \hat{G}_\delta(h,x) - \psi_n(I;x)x + q_n(x)x^2,
\end{aligned} \tag{2.92}$$

where $q_n(x) := \frac{1}{2}\frac{\partial^2 \log P\big(Y_n \in [y,y+\delta] \mid X_n = x\big)}{\partial y^2}\Big|_{y=h-\hat{\alpha}_n x}$. Therefore, for the second term in (2.83), by (2.92), we can get

$$\log\left(\frac{P\big(Y_n \in [h-x, h+\delta-x] \mid X_n = x\big)}{P\big(Y_n \in [h,h+\delta] \mid X_n = x\big)e^{-\psi_n(I;x)x}}\right) = \hat{G}_\delta(h-x,x) - \hat{G}_\delta(h,x) + \psi_n(I;x)x = q_n(x)x^2.$$

And by Condition (2.36), for all $x \in \mathbb{R}^+$, there is a constant $K_2 > 0$ such that

$$|e^{-\psi_n(I;x)x} q_n(x)| \le K_2. \tag{2.93}$$

In the following, we use a brief notation $P_{Y_n|X_n}\big(E_n - x; x\big) = P\big(Y_n \in [y, y+\delta] \mid X_n = x\big)$. By (2.93), and the uniform boundedness of A_n, and the assumption: $\mathbb{E}\big[X_n^2\big] = a_n$, the second term in (2.83) satisfies

$$\begin{aligned}
&\left|\int_{\mathbb{R}^+} A_n f_{X_n}(x) e^{-\psi_n(I;x)x} \log\left(\frac{P_{Y_n|X_n}\big(I - x; x\big)}{P_{Y_n|X_n}\big(I; x\big)e^{-\psi_n(I;x)x}}\right) dx\right| \\
=&\left|\int_{\mathbb{R}^+} A_n e^{-\psi_n(I;x)x} f_{X_n}(x) q_n(x) x^2 dx\right| \le M\left|\int_{\mathbb{R}^+} f_{X_n}(x) e^{-\psi_n(I;x)x} q_n(x) x^2 dx\right| \le M K_2 \mathbb{E}\big[X_n^2\big] = O(a_n).
\end{aligned} \tag{2.94}$$

Combining (2.82), (2.83), (2.91) and (2.94),

$$D_{\mathrm{KL}}\left(\mathbb{P}_I^{(n)} \parallel \mathbb{Q}_I^{(n)}\right) = O(a_n + b_n). \tag{2.95}$$

For the case $S_n := \{x : f_{X_n}(x) > 0\} \subset \mathbb{R}^+$, we can only define $P(Z_n \in I \mid X_n = x)$ on S_n. But we can still define the KL-divergence on $\mathbb{R}^+$ since the part of KL-divergence on $\mathbb{R}^+ \backslash S_n$ is 0. Therefore, in the same way as (2.81),

$$\begin{aligned}
D_{\mathrm{KL}}\left(\mathbb{P}_I^{(n)} \parallel \mathbb{Q}_I^{(n)}\right) &= \int_{\mathbb{R}^+} A_n f_{X_n}(x) e^{-\psi_n(I;x)x} \log\left(\frac{A_n f_{X_n}(x) e^{-\psi_n(I;x)x}}{f_{X_n|Z_n}(x, I)}\right) dx \\
&= \int_{S_n} A_n f_{X_n}(x) e^{-\psi_n(I;x)x} \log\left(\frac{A_n f_{X_n}(x) e^{-\psi_n(I;x)x}}{f_{X_n|Z_n}(x, I)}\right) dx \\
&= \left|\int_{S_n} A_n f_{X_n}(x) e^{-\psi_n(I;x)x} \log\left(\frac{f_{X_n|Z_n}(x, I)}{A_n f_{X_n}(x) e^{-\psi_n(I;x)x}}\right) dx\right|.
\end{aligned} \tag{2.96}$$

Then we can follow every step from the step (2.82) in our proof for the case $\{x : f_{X_n}(x) > 0\} = \mathbb{R}^+$ to get

$$D_{\mathrm{KL}}\left(\mathbb{P}_I^{(n)} \parallel \mathbb{Q}_I^{(n)}\right) = O(a_n + b_n). \tag{2.97}$$

Furthermore, let $\zeta_n(I; x) := \beta_n \psi_n(I; \beta_n x)$, by the condition (2.37), there is a constant $C > 0$ such that for all $x \in \mathbb{R}^+$, $0 \le \zeta_n(I; x) < C$. Therefore, $\mathbb{P}_I^{(n)}$ with the density function $A_n f_X(x) e^{-\beta_n \psi_n(I;\beta_n x)x}$ satisfies the definition of the canonical probability distributions in (2.31). $\square$

Proof of Theorem 2.3.6

For a finite interval $I = [h, h+\delta],\ h, \delta \in \mathbb{R}$ and $\delta > 0$, let

$$\hat{\mathbb{P}}_I^{(n)} = P(K_n = \beta_n k \mid Z_n \in I) \quad \text{and} \quad \hat{\mathbb{Q}}_I^{(n)} = B_n P(K_n = \beta_n k) e^{-\hat{\psi}(I;\beta_n k)\beta_n k},$$

where

$$\frac{1}{B_n} := \sum_{\beta_n k \in S_n} P(K_n = \beta_n k) e^{-\hat{\psi}(I;\beta_n k)\beta_n k}$$

and

$$\hat{\psi}_n(I;\beta_n k) := \left. \frac{\partial \log P\big(Y_n \in [y, y+\delta] \mid K_n = \beta_n k\big)}{\partial y} \right|_{y=h}.$$

We first state the following lemma. The proof follows from the proof of Theorem 2.3.1 with the Definition of KL-divergence for discrete probability distributions in (2.21).

Lemma 2.4.1. *Assume there exist positive constants $\delta_1, \delta_2, \{C_i, 1 \leq i \leq 3\}$, a sequence $b_n = o(1)$, and an open interval D such that the conditions* (2.36) – (2.39) *in Theorem 2.3.1 hold for K_n, Y_n, and Z_n. Then*

$$D_{\text{KL}}\left(\hat{\mathbb{P}}_I^{(n)} \| \, \hat{\mathbb{Q}}_I^{(n)}\right) = O(a_n + b_n), \quad \textit{for every } I \subseteq D. \tag{2.98}$$

Now we are ready to prove Theorem 2.3.6.

Proof of Theorem 2.3.6. All of the conditions in Theorem 2.3.1 hold for K_n, Y_n, Z_n by the assumptions, hence Lemma 2.4.1 can be applied. Therefore, we obtain the following relation between total variation and KL-divergence from (2.22): for every $I \subseteq D$,

$$\begin{aligned} &\sup_{\beta_n k \in S_n} \left| P\big(K_n = \beta_n k \mid Z_n \in I\big) - B_n P(K_n = \beta_n k) e^{-\hat{\psi}_n(I;\beta_n k)\beta_n k} \right| \\ \leq &\delta\left(\hat{\mathbb{P}}_I^{(n)}, \hat{\mathbb{Q}}_I^{(n)}\right) \leq \frac{1}{2}\sqrt{D_{\text{KL}}\left(\hat{\mathbb{P}}_I^{(n)} \| \, \hat{\mathbb{Q}}_I^{(n)}\right)} = O\left(\sqrt{a_n + b_n}\right). \end{aligned} \tag{2.99}$$

With (2.52) and (2.99), the conclusion (2.53) follows from the change of variable $K_n = \beta_n K$ and the triangle inequality:

$$
\begin{aligned}
&\sup_{k\in S}\left|P\big(K=k \mid \tilde{H}_n \in E_n\big) - B_n P(K=k)e^{-\beta_n\hat{\psi}_n(I;\beta_n k)k}\right| \\
=&\sup_{\beta_n k\in S_n}\left|P\big(K_n=\beta_n k \mid H_n \in I\big) - B_n P(K_n=\beta_n k)e^{-\hat{\psi}_n(I;\beta_n k)\beta_n k}\right| \\
\leq&\sup_{\beta_n k\in S_n}\left|P\big(K_n=\beta_n k \mid Z_n \in I\big) - B_n P(K_n=\beta_n k)e^{-\hat{\psi}_n(I;\beta_n k)\beta_n k}\right| \\
&+\sup_{\beta_n k\in S_n}\left|P\big(K_n=\beta_n k \mid H_n \in I\big) - P\big(K_n=\beta_n k \mid Z_n \in I\big)\right| \\
=&O(c_n+\sqrt{a_n+b_n}).
\end{aligned}
$$

Furthermore, let $\hat{\zeta}_n(I;k) := \beta_n\hat{\psi}_n(I;\beta_n k)$. We can check that $0 \leq \hat{\zeta}_n(I;k) < C$ for all $k \in S$ and a constant $C > 0$. Therefore, $B_n P(K=k)e^{-\beta_n\hat{\psi}_n(I;\beta_n k)k}$ satisfies the definition of the canonical probability distributions in (2.32). □

2.4.2 Proofs of Theorem 2.3.8 and Theorem 2.3.9

Let X be a nonnegative continuous random variable and with $\mathbb{E}[X] < \infty$ and let Z_n be a sequence of real-valued continuous random variables. Given a Borel measurable set $E \in \mathcal{B}(\mathbb{R})$ and a function $\psi : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$ with $0 < \psi(E) < \infty$, let $\mathbb{P}_E$ be a probability measure with density function

$$
Af_X(x)e^{-\psi(E)x}, \quad \frac{1}{A} := \int_{\mathbb{R}^+} f_X(x)e^{-\psi(E)x}dx.
$$

And let $\mathbb{Q}_E^{(n)}$ be a probability measure with density function $f_{X|Z_n}(x;E)$. We obtain the following lemma for the case (2) of the canonical distribution (2.58):

Lemma 2.4.2. *Assume the following conditions hold:*

1. *(Boundedness)* $\left|\dfrac{f_{X|Z_n}(x;E)}{f_X(x)}\right|$ *and* $\left|e^{-\xi x}\log\left(\dfrac{f_{X|Z_n}(x;E)}{f_X(x)}\right)\right|$*, for any $\xi > 0$, are uniformly bounded on $\mathbb{R}^+$.*

2. *(Linear approximation) There exist constants* $b, c \in \mathbb{R},\ 0 < c < \infty$*, and a sequence of functions* $q_n : \mathbb{R} \to \mathbb{R}$ *with*

$$\mathbb{E}\left[q_n(X)^2\right] = \gamma_n \to 0$$

such that on an interval $I_n = [0, d_n]$ *with* $d_n \to \infty$*,*

$$\log\left(\frac{f_{X|Z_n}(x;E)}{f_X(x)}\right) = b - cx + q_n(x). \tag{2.100}$$

Then

$$\lim_{n\to\infty} D_{\mathrm{KL}}\left(\mathbb{P}_E \parallel \mathbb{Q}_E^{(n)}\right) = 0 \quad \textit{if and only if} \quad c = \psi(E).$$

Furthermore, assume $\mathbb{E}[X^3] < \infty$ and X is not a constant random variable, let $\tilde{\mathbb{P}}_E^{(n)}$ be a probability measure with density function

$$\tilde{A}_n f_X(x) e^{-\beta_n \psi(E) x}, \qquad \frac{1}{\tilde{A}_n} := \int_{\mathbb{R}^+} f_X(x) e^{-\beta_n \psi(E) x}, \tag{2.101}$$

in which $\beta_n > 0,\ \beta_n = o(1)$. We obtain the following lemma for the case (1) of the canonical distribution (2.58):

Lemma 2.4.3. *Assume the following conditions hold :*

1. *(Boundedness)* $\left|\frac{f_{X|Z_n}(x;E)}{f_X(x)}\right|$ *and* $\left|e^{-\beta_n \xi x} \log\left(\frac{f_{X|Z_n}(x;E)}{f_X(x)}\right)\right|$*, for any* $\xi > 0$*, are uniformly bounded on* $\mathbb{R}^+$*.*

2. *(Linear approximation) There exist constants* $b, c \in \mathbb{R},\ 0 < c < \infty$*, and a sequence of functions* $q_n : \mathbb{R} \to \mathbb{R}$ *with* $\mathbb{E}\left[q_n(X)^2\right] \to 0$ *such that on* $I_n = [0, d_n]$ *with* $d_n = O\left(\frac{1}{\beta_n}\right)$*,*

$$\frac{1}{\beta_n} \log\left(\frac{f_{X|Z_n}(x;E)}{f_X(x)}\right) = b - cx + q_n(x). \tag{2.102}$$

Then

$$\lim_{n\to\infty} \frac{D_{\mathrm{KL}}\left(\tilde{\mathbb{P}}_E^{(n)} \parallel \mathbb{Q}_E^{(n)}\right)}{\beta_n^2} = 0 \quad \textit{if and only if} \quad c = \psi(E).$$

Corollary 2.4.4. *In particular, if we choose* $Z_n = \beta_n X + \beta_n(\tilde{Y}_n - \mu_n)$*, where* $\tilde{Y}_n, \beta_n, \mu_n$ *are given in the definitions in Section 2.3.2, and choose the Borel set* E *to be a finite interval* I*, by Equation* (2.27)*, those general results of Lemma 2.4.2 and Lemma 2.4.3 for* $f_{X|Z_n}(x, E)$ *can be applied to* $f_{X|\tilde{Z}_n}(x, E_n)$*, which is the conditional density defined in Section 2.3.2.*

Proof of Lemma 2.4.2

Proof. Note that for any uniformly bounded function $|b_n(x)|$ on $\mathbb{R}^+$:

$$\begin{aligned}\left|\int_{\mathbb{R}^+\setminus I_n} f_X(x)b_n(x)dx\right| &\leq \|b_n(x)\|_\infty \int_{\mathbb{R}^+\setminus I_n} f_X(x)dx = \|b_n(x)\|_\infty \mathbb{P}\left(X \geq d_n\right) \\ &\leq \|b_n(x)\|_\infty \left(\frac{\mathbb{E}\left[X\right]}{d_n}\right) = O(\epsilon_n),\end{aligned} \tag{2.103}$$

for a sequence $\epsilon_n \to 0$ since $d_n \to \infty$ by Condition (2) and $\mathbb{E}[X]$ is bounded by the assumption.

We first prove $c = \psi(E) \;\Rightarrow\; D_{\text{KL}}\left(\mathbb{P}_E \parallel \mathbb{Q}_E^{(n)}\right) \to 0$.

By Condition (2),

$$\log\left(\frac{f_{X|Z_n}(x;E)}{f_X(x)}\right) = b - \psi(E)x + q_n(x) \quad \text{on } I_n, \tag{2.104}$$

therefore, we have

$$\log\left(\frac{Af_X(x)e^{-\psi(E)x}}{f_{X|Z_n}(x;E)}\right) = \log A - b - q_n(x) \quad \text{on } I_n. \tag{2.105}$$

Since $\int_{\mathbb{R}^+} f_X(x)dx = 1$, there exists a bounded closed set $D \subset \mathbb{R}^+$ such that $\int_D f_X(x)dx > 0$. Hence,

$$A = \frac{1}{\displaystyle\int_{\mathbb{R}^+} f_X(x)e^{-\psi(E)x}dx} \leq \frac{1}{\displaystyle\int_D f_X(x)e^{-\psi(E)x}dx} \leq \frac{1}{\displaystyle\inf_{x\in D}\left|e^{-\psi(E)x}\right|\int_D f_X(x)dx} < \infty. \tag{2.106}$$

Furthermore, we can derive

$$\begin{aligned}1 = \int_{\mathbb{R}^+} f_{X|Z_n}(x;E)dx &= \int_{I_n} f_X(x)\frac{f_{X|Z_n}(x;E)}{f_X(x)}dx + \int_{\mathbb{R}^+\setminus I_n} f_X(x)\frac{f_{X|Z_n}(x;E)}{f_X(x)}dx \\ &= \int_{I_n} f_X(x)e^{b-\psi(E)x+q_n(x)}dx + O(\epsilon_n),\end{aligned} \tag{2.107}$$

in which the last equality is from Equation (2.104), and the result of (2.103) applied to the uniformly bounded function $|b_n(x)| = \left|\dfrac{f_{X|Z_n}(x;E)}{f_X(x)}\right|$ on $\mathbb{R}^+$ (Condition (1)). Multiplying by e^{-b} on both sides in (2.107), we have

$$e^{-b} = \int_{I_n} f_X(x)e^{-\psi(E)x+q_n(x)}dx + O(\epsilon_n). \tag{2.108}$$

Then we can apply Taylor's expansion to $e^{q_n(x)}$ to get

$$e^{-b} = \int_{I_n} f_X(x) e^{-\psi(E)x} dx + \int_{I_n} f_X(x) e^{-\psi(E)x} \left(q_n(x) + \frac{q_n(x)^2}{2} e^{\alpha_n \cdot q_n(x)} \right) dx + O(\epsilon_n), \quad (2.109)$$

for some sequence $\alpha_n \in (0,1)$. Note that we use the formula

$$e^y = 1 + y + \frac{e^{\alpha(y)\cdot y}}{2} y^2, \quad \alpha(y) \in (0,1)$$

and let $y = q_n(x),\ \alpha_n = \alpha(q_n(x))$. Combined with Equation (2.104) and Condition (1), it implies there exists a constant $M > 1$ independent of n such that $e^{-\psi(E)x + q_n(x)} \leq M$ for all $x \in I_n$. Since $\alpha_n \in (0,1)$ and $\psi(E) > 0$ in the assumption,

$$e^{-\psi(E)x + \alpha_n \cdot q_n(x)} \leq M^{\alpha_n} e^{-(1-\alpha_n)\psi(E)x} \leq M \quad \text{for all } x \in I_n. \quad (2.110)$$

Hence $e^{-\psi(E)x + \alpha_n q_n(x)}$ is uniformly bounded on I_n. Then

$$\begin{aligned} \int_{I_n} f_X(x) e^{-\psi(E)x} \left(\frac{q_n(x)^2}{2} e^{\alpha_n \cdot q_n(x)} \right) dx &\leq \left\| \frac{e^{-\psi(E)x + \alpha_n \cdot q_n(x)}}{2} \right\|_\infty \int_{I_n} f_X(x) q_n(x)^2 dx \\ &\leq M \mathbb{E}\left[q_n(X)^2 \right] = O(\gamma_n), \end{aligned} \quad (2.111)$$

where $O(\gamma_n) \to 0$ by Condition (2). By Equations (2.109) and (2.111),

$$\begin{aligned} e^{-b} &\leq \int_{I_n} f_X(x) e^{-\psi(E)x} dx + \int_{I_n} f_X(x) e^{-\psi(E)x} q_n(x) dx + O(\gamma_n) + O(\epsilon_n) \\ &= \int_{\mathbb{R}^+} f_X(x) e^{-\psi(E)x} dx - \int_{\mathbb{R}^+ \setminus I_n} f_X(x) e^{-\psi(E)x} dx + \int_{I_n} f_X(x) e^{-\psi(E)x} q_n(x) dx + O(\gamma_n) + O(\epsilon_n) \\ &= \frac{1}{A} + \int_{I_n} f_X(x) e^{-\psi(E)x} q_n(x) dx + O(\gamma_n) + O(\epsilon_n), \end{aligned} \quad (2.112)$$

where the last equation is from the result of (2.103). And since A is bounded by the result (2.106), we have

$$A e^{-b} \leq 1 + \int_{I_n} A f_X(x) e^{-\psi(E)x} q_n(x) dx + O(\gamma_n) + O(\epsilon_n). \quad (2.113)$$

Using the inequality $\log(1+x) \leq x$ for all $x > -1$, we find a bound

$$\log A - b \leq \int_{I_n} A f_X(x) e^{-\psi(E)x} q_n(x) dx + O(\gamma_n) + O(\epsilon_n). \quad (2.114)$$

Furthermore, by Condition (1), $\left|e^{-\psi(E)x}\log\left(\frac{f_{X|Z_n}(x;E)}{f_X(x)}\right)\right|$ is uniformly bounded on $\mathbb{R}^+$, so we can check that

$$\left|e^{-\psi(E)x}\log\left(\frac{Af_X(x)e^{-\psi(E)x}}{f_{X|Z_n}(x;E)}\right)\right| \text{ is uniformly bounded on } \mathbb{R}^+ \text{as well.} \tag{2.115}$$

Recall that

$$\log\left(\frac{Af_X(x)e^{-\psi(E)x}}{f_{X|Z_n}(x;E)}\right) = \log A - b - q_n(x) \text{ on } I_n$$

by (2.105). With the result of (2.114), we can get

$$\begin{aligned}
&D_{\text{KL}}\left(\mathbb{P}_E \parallel \mathbb{Q}_E^{(n)}\right)\\
&= \int_{I_n} Af_X(x)e^{-\psi(E)x}\log\left(\frac{Af_X(x)e^{-\psi(E)x}}{f_{X|Z_n}(x;E)}\right)dx + \int_{\mathbb{R}^+\setminus I_n} Af_X(x)e^{-\psi(E)x}\log\left(\frac{Af_X(x)e^{-\psi(E)x}}{f_{X|Z_n}(x;E)}\right)dx\\
&= \int_{I_n} Af_X(x)e^{-\psi(E)x}\left(\log A - b\right)dx - \int_{I_n} Af_X(x)e^{-\psi(E)x}q_n(x)dx + O(\epsilon_n)\\
&= (\log A - b) - \int_{\mathbb{R}^+\setminus I_n} Af_X(x)e^{-\psi(E)x}\left(\log A - b\right)dx - \int_{I_n} Af_X(x)e^{-\psi(E)x}q_n(x)dx + O(\epsilon_n)\\
&= \log A - b + O(\epsilon_n) - \int_{I_n} Af_X(x)e^{-\psi(E)x}q_n(x)dx + O(\epsilon_n) = O(\gamma_n) + O(\epsilon_n), \qquad (2.116)
\end{aligned}$$

where the $O(\epsilon_n)$ terms are from the result of (2.103) applied to the bounded function (2.115). Therefore, by (2.106) and (2.116), we get

$$D_{\text{KL}}\left(\mathbb{P}_E \parallel \mathbb{Q}_E^{(n)}\right) \to 0.$$

Next we prove

$$D_{\text{KL}}\left(\mathbb{P}_E \parallel \mathbb{Q}_E^{(n)}\right) \to 0 \;\Rightarrow\; c = \psi(E). \tag{2.117}$$

By Condition (2), there exists a constant $\hat{b}$ and a sequence of functions $\hat{q}_n(x)$ such that

$$\log\left(\frac{f_{X|Z_n}(x;E)}{f_X(x)}\right) = b - cx + q_n(x) \quad \text{on } I_n. \tag{2.118}$$

Similar to the derivation of (2.105), we have

$$\log\left(\frac{\hat{A}f_X(x)e^{-cx}}{f_{X|Z_n}(x;E)}\right) = \log\hat{A} - b - q_n(x) \quad \text{on } I_n, \tag{2.119}$$

where

$$\hat{A} = \frac{1}{\displaystyle\int_{\mathbb{R}^+} f_X(x)e^{-cx}dx} < \infty, \tag{2.120}$$

which can be proved by a similar approach as in (2.106). Then following the previous proof from (2.107) to (2.116), we can get

$$D_{\text{KL}}\left(\hat{\mathbb{P}}_E \parallel \mathbb{Q}_E^{(n)}\right) \to 0, \tag{2.121}$$

where $\hat{\mathbb{P}}_E$ is a probability measure with density function $\hat{A}f_X(x)e^{-cx}$. By the assumption (2.117), we also know

$$D_{\text{KL}}\left(\mathbb{P}_E \parallel \mathbb{Q}_E^{(n)}\right) \to 0. \tag{2.122}$$

By Pinsker's inequality (2.22), we have that the total variation distance denoted by $\delta(\cdot,\cdot)$ satisfies

$$\delta(\hat{\mathbb{P}}_E, \mathbb{Q}_E^{(n)}) \to 0 \quad \text{and} \quad \delta(\mathbb{P}_E, \mathbb{Q}_E^{(n)}) \to 0. \tag{2.123}$$

Then by the triangle inequality, $\delta(\hat{\mathbb{P}}_E, \mathbb{P}_E) = 0$. It implies

$$\int_0^x \left(\hat{A}f_X(s)e^{-cs}ds - Af_X(s)e^{-\psi(E)s}\right) ds = 0, \qquad \text{for all } x \in \mathbb{R}^+.$$

Hence

$$\hat{A}f_X(x)e^{-cx} = Af_X(x)e^{-\psi(E)x} \quad \text{holds almost everywhere on } \mathbb{R}^+.$$

Since $\hat{A}$ and A are both independent of x and there exists an interval such that $f_X(x) > 0$, we obtain $c = \psi(E)$. □

Proof of Lemma 2.4.3

Proof. Note that for any uniform bounded function $|b_n(x)|$ on $\mathbb{R}^+$:

$$\begin{aligned} \left|\int_{\mathbb{R}^+\setminus I_n} f_X(x)b_n(x)dx\right| &\le \|b_n(x)\|_\infty \int_{\mathbb{R}^+\setminus I_n} f_X(x)dx = \|b_n(x)\|_\infty \mathbb{P}\left(X \ge d_n\right) \\ &\le \|b_n(x)\|_\infty \left(\frac{\mathbb{E}\left[X^3\right]}{d_n^3}\right) = O(\beta_n^3), \end{aligned} \tag{2.124}$$

where the existence of $O(\beta_n^3)$ is due to $d_n = O(\frac{1}{\beta_n})$ by Condition (2) and $\mathbb{E}[X^3] < \infty$ by the assumption.

We first prove $c = \psi(E) \Rightarrow \dfrac{D_{\text{KL}}\left(\tilde{\mathbb{P}}_E^{(n)} \parallel \mathbb{Q}_E^{(n)}\right)}{\beta_n^2} \to 0$. By Condition (2),

$$\log\left(\frac{f_{X|Z_n}(x;E)}{f_X(x)}\right) = \beta_n\left(b - \psi(E)x + q_n(x)\right) \quad \text{on } I_n, \tag{2.125}$$

Therefore we have

$$\log\left(\frac{A_n f_X(x) e^{-\beta_n \psi(E)x}}{f_{X|Z_n}(x;E)}\right) = \log A_n - \beta_n b - \beta_n q_n(x) \quad \text{on } I_n. \tag{2.126}$$

Following the proof in (2.106), for each n, we have

$$A_n = \frac{1}{\displaystyle\int_{\mathbb{R}^+} f_X(x) e^{-\beta_n \psi(E)x} dx} < \infty, \tag{2.127}$$

and we can check that $\lim\limits_{n\to\infty} \displaystyle\int_{\mathbb{R}^+} f_X(x) e^{-\beta_n\psi(E)x} dx \to 1$, hence, A_n is uniformly bounded.

We can apply a similar proof as for Lemma 2.4.2 to Equation (2.126). Substituting b by $\beta_n b$, $\psi(E)x$ by $\beta_n \psi(E)x$, $q_n(x)$ by $\beta_n q_n(x)$ and A by A_n, then every step from Equation (2.107) to Equation (2.116) follows. Therefore, we can get

$$D_{\text{KL}}\left(\tilde{\mathbb{P}}_E^{(n)} \parallel \mathbb{Q}_E^{(n)}\right) = O(\beta_n^2 \gamma_n) + O(\beta_n^3),$$

where the $O(\beta_n^2\gamma_n)$ term follows from the derivation of the $O(\gamma_n)$ term in Lemma 2.4.2, the $O(\beta_n^3)$ term follows from Equation (2.124) and the derivation of the $O(\epsilon_n)$ term in Lemma 2.4.2. It implies

$$\frac{D_{\text{KL}}\left(\tilde{\mathbb{P}}_E^{(n)} \parallel \mathbb{Q}_E^{(n)}\right)}{\beta_n^2} = O(\gamma_n) + O(\beta_n) \to 0.$$

Next we prove $\dfrac{D_{\text{KL}}\left(\tilde{\mathbb{P}}_E^{(n)} \parallel \mathbb{Q}_E^{(n)}\right)}{\beta_n^2} \to 0 \Rightarrow c = \psi(E)$. Similar to the proof for Lemma 2.4.2, we can show

$$\frac{D_{\text{KL}}\left(\hat{\mathbb{P}}_E^{(n)} \parallel \mathbb{Q}_E^{(n)}\right)}{\beta_n^2} \to 0, \tag{2.128}$$

where $\hat{\mathbb{P}}_E^{(n)}$ is a probability measure with density function $\hat{A}_n f_X e^{-\beta_n c x}$. By the assumption, we also know

$$\frac{D_{\text{KL}}\left(\tilde{\mathbb{P}}_E^{(n)} \parallel \mathbb{Q}_E^{(n)}\right)}{\beta_n^2} \to 0. \tag{2.129}$$

Therefore, by Pinsker's inequality, we have that the total variation distance denoted by $\delta(\cdot,\cdot)$ satisfies

$$\frac{1}{\beta_n}\delta(\hat{\mathbb{P}}_E, \mathbb{Q}_E^{(n)}) \to 0 \quad \text{and} \quad \frac{1}{\beta_n}\delta(\tilde{\mathbb{P}}_E, \mathbb{Q}_E^{(n)}) \to 0. \tag{2.130}$$

Then by the triangle inequality, $\frac{1}{\beta_n}\delta(\hat{\mathbb{P}}_E, \tilde{\mathbb{P}}_E) \to 0$. It implies

$$\lim_{n\to\infty} \frac{1}{\beta_n}\left(\int_0^x \hat{A}_n f_X(s) e^{-\beta_n c s} ds - \int_0^x A_n f_X(s) e^{-\beta_n \psi(E) s} ds\right) = 0, \quad \text{for all } x \in \mathbb{R}^+. \tag{2.131}$$

We can apply Taylor's expansion to $e^{-\beta_n c s}$ and $e^{-\beta_n \psi(E) s}$ to get

$$e^{-\beta_n c s} = 1 - \beta_n c s + O(\beta_n^2 s^2) \quad \text{and} \quad e^{-\beta_n \psi(E) s} = 1 - \beta_n \psi(E) s + O(\beta_n^2 s^2). \tag{2.132}$$

By the results of (2.132), Equation (2.131) can be written as

$$\lim_{n\to\infty} \int_0^x \frac{1}{\beta_n}\left(\tilde{A}_n - A_n - \left(\tilde{A}_n c - A_n \psi(E)\right)\beta_n s + O(\beta_n^2 s^2)\right) f_X(s) ds = 0, \quad \text{for all } x \in \mathbb{R}^+. \tag{2.133}$$

Since $\mathbb{E}[X^2] < \infty$ from the fact $\mathbb{E}[X^3] < \infty$, we know $\displaystyle\int_0^x s^2 f_X(s) ds < \infty$ on $\mathbb{R}^+$. Therefore, the $O(\beta_n^2 s^2)$ in Equation (2.133) can be dropped and we obtain

$$\lim_{n\to\infty} \int_0^x \frac{1}{\beta_n}\left(\tilde{A}_n - A_n - \left(\tilde{A}_n c - A_n \psi(E)\right)\beta_n s\right) f_X(s) ds = 0, \quad \text{for all } x \in \mathbb{R}^+. \tag{2.134}$$

By the Dominated Convergence Theorem,

$$\hat{A}_n = \frac{1}{\int_{\mathbb{R}^+} f_X(x) e^{-\beta_n c x} dx} \to 1 \quad \text{and} \quad A_n = \frac{1}{\int_{\mathbb{R}^+} f_X(x) e^{-\beta_n \psi_n(E) x} dx} \to 1. \tag{2.135}$$

Also we have

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{\beta_n}\left(\int_{\mathbb{R}^+} f_X(x) e^{-\beta_n \psi_n(E) x} dx - \int_{\mathbb{R}^+} f_X(x) e^{-\beta_n c x} dx\right) \\
&= \lim_{n\to\infty} \frac{1}{\beta_n}\left(\int_{\mathbb{R}^+} \left((c - \psi(E))\,\beta_n x + O(\beta_n^2 x^2)\right) f_X(x) dx\right) \\
&= (c - \psi(E))\,\mathbb{E}[X] + \lim_{n\to\infty} O(\beta_n \mathbb{E}[X^2]) = (c - \psi(E))\,\mathbb{E}[X],
\end{aligned} \tag{2.136}$$

where in the first equality we apply Taylor's expansion (2.132) again. By (2.135) and (2.136), we have

$$\lim_{n\to\infty}\frac{\tilde{A}_n - A_n}{\beta_n} = \lim_{n\to\infty}\frac{1}{\beta_n}\left(\frac{\int_{\mathbb{R}^+} f_X(x)e^{-\beta_n\psi_n(E)x}dx - \int_{\mathbb{R}^+} f_X(x)e^{-\beta_n cx}dx}{\int_{\mathbb{R}^+} f_X(x)e^{-\beta_n cx}dx\int_{\mathbb{R}^+} f_X(x)e^{-\beta_n\psi_n(E)x}dx}\right) = (c-\psi(E))\,\mathbb{E}[X]. \tag{2.137}$$

Therefore from (2.134) and (2.137),

$$\lim_{n\to\infty}\int_0^x \left(\tilde{A}_n c - A_n\psi(E)\right) sf_X(s)ds = (c-\psi(E))\int_0^x sf_X(s)ds, \quad \text{for all } x\in\mathbb{R}^+. \tag{2.138}$$

Therefore, we can apply the results of (2.137) and (2.138) to Equation (2.134) to get

$$(c-\psi(E))\int_0^x \mathbb{E}[X]f_X(s)ds = (c-\psi(E))\int_0^x sf_X(s)ds, \quad \text{for all } x\in\mathbb{R}^+. \tag{2.139}$$

Since X is is not a constant random variable by our assumption, (2.139) is only true when $c = \psi(E)$. ☐

Proof of Theorem 2.3.8

Proof. The proof follows from Lemma 2.4.3. By the condition (2) in Theorem 2.3.8, we have

$$\begin{aligned}\log\left(\frac{P\left(Y_n\in I-\beta_n x \mid X_n=\beta_n x\right)}{P\left(Z_n\in I\right)}\right) &= \log\left(\frac{P\left(Y\in I-\beta_n x\right)}{P\left(Y\in I\right)}\right) + g_n(x)\\ &= -\left.\frac{\partial \log P(Y\in[y,y+\delta])}{\partial y}\right|_h \beta_n x + O(\beta_n^2x^2) + \beta_n g_n(x).\end{aligned} \tag{2.140}$$

We now check whether all conditions in Lemma 2.4.3 are satisfied:

1. (Boundedness): $\left|\frac{f_{X|Z_n}(x;I)}{f_X(x)}\right| = \left|\frac{f_{X|\tilde{Z}_n}(x;E_n)}{f_X(x)}\right|$, which is uniformly bounded on $\mathbb{R}^+$ by the condition (2) in Theorem 2.3.8. And from (2.140), for any $\xi>0$,

$$\begin{aligned}&\left|e^{-\beta_n\xi x}\log\left(\frac{f_{X|Z_n}(x;I)}{f_X(x)}\right)\right| = \left|e^{-\beta_n\xi x}\log\left(\frac{P\left(Y_n\in I-\beta_n x\mid X_n=\beta_n x\right)}{P\left(Z_n\in I\right)}\right)\right|\\ &\le \left|e^{-\beta_n\xi x}O(\beta_n x+\beta_n^2x^2)\right| + \left|e^{-\beta_n\xi x}g_n(x)\right|,\end{aligned}$$

where the first term is uniformly bouneded on $\mathbb{R}^+$, and the second term is uniformly bouneded on $\mathbb{R}^+$ by the condition (3) in Theorem 2.3.8.

2. (Linear approximation): Following (2.140), we have

$$\begin{aligned}\frac{1}{\beta_n}\log\left(\frac{f_{X|Z_n}(x;E)}{f_X(x)}\right) &= \frac{1}{\beta_n}\log\left(\frac{P\left(Y_n \in I-\beta_n x \mid X_n=\beta_n x\right)}{P\left(Z_n \in I\right)}\right)\\ &= -\left.\frac{\partial \log P(Y\in[y,y+\delta])}{\partial y}\right|_h x + O(\beta_n x^2) + g_n(x)\end{aligned}$$

on $I_n=[0,d_n]$ with $d_n = O\left(\frac{1}{\beta_n}\right)$. Therefore, we obtain

$$c = \left.\frac{\partial \log P(Y\in[y,y+\delta])}{\partial y}\right|_h \quad \text{and} \quad q_n(x) = O(\beta_n x^2) + g_n(x)$$

and we can check that $\mathbb{E}[q_n(X)^2]\to 0$ since $\mathbb{E}[g_n(X)^2]\to 0$ by the condition (3).

Therefore, applying Lemma 2.4.3, we have

$$\lim_{n\to\infty}\frac{D_{\mathrm{KL}}\left(\tilde{\mathbb{P}}_I^{(n)} \parallel \mathbb{Q}_I^{(n)}\right)}{\beta_n^2} = 0 \quad \text{if and only if} \quad \psi(I) = \left.\frac{\partial \log P(Y\in[y,y+\delta])}{\partial y}\right|_h.$$

Furthermore, since $0 < \left.\frac{\partial \log P(Y\in[y,y+\delta])}{\partial y}\right|_h < C$ for a constant $C>0$, $\tilde{\mathbb{P}}_I^{(n)}$ satisfies the definition of the canonical probability distributions in (2.31).

$\square$

Proof of Theorem 2.3.9

Proof. The proof follows from Lemma 2.4.2. By the condition (2) in Theorem 2.3.9, we have

$$\begin{aligned}\log\left(\frac{P\left(Y_n\in I-\beta_n x \mid X_n=\beta_n x\right)}{P\left(Z_n\in I\right)}\right) &= \log\left(\frac{\exp\left[-\frac{1}{\beta_n}\phi\left(y^*-\beta_n x\right)\right]}{\exp\left[-\frac{1}{\beta_n}\phi\left(y^*\right)\right]}\right) + r_n(x)\\ &= \phi'(y^*)x + O(\beta_n x^2) + r_n(x). \qquad (2.141)\end{aligned}$$

To check that all conditions are satisfied:

1. (Boundedness): $\left|\frac{f_{X|Z_n}(x;I)}{f_X(x)}\right| = \left|\frac{f_{X|\tilde{Z}_n}(x;E_n)}{f_X(x)}\right|$, which is uniformly bounded on $\mathbb{R}^+$ by the condition (1) in Theorem 2.3.9. And by (2.141), for any $\xi > 0$,

$$\left|e^{-\xi x}\log\left(\frac{f_{X|Z_n}(x;I)}{f_X(x)}\right)\right| = \left|e^{-\xi x}\log\left(\frac{P\left(Y_n \in I - \beta_n x \mid X_n = \beta_n x\right)}{P\left(Z_n \in I\right)}\right)\right|$$
$$\leq \left|e^{-\xi x} O(x + \beta_n x^2)\right| + \left|e^{-\xi x} r_n(x)\right|,$$

where the first term is uniformly bounded on $\mathbb{R}^+$, and the second term is uniformly bounded on $\mathbb{R}^+$ by the condition (3) in Theorem 2.3.9.

2. (Linear approximation): As follows from (2.141), we have

$$\log\left(\frac{f_{X|Z_n}(x;E)}{f_X(x)}\right) = \log\left(\frac{P\left(Y_n \in I - \beta_n x \mid X_n = \beta_n x\right)}{P\left(Z_n \in I\right)}\right)$$
$$= \phi'(y^*)x + O(\beta_n x^2) + r_n(x)$$

on $I_n = [0, d_n]$ with $d_n = O\left(\frac{1}{\beta_n}\right)$. Hence we obtain

$$c = -\phi'(y^*) \quad \text{and} \quad q_n(x) = O(\beta_n x^2) + r_n(x),$$

and we can check that $\mathbb{E}[q_n(X)^2] \to 0$ since $\mathbb{E}[r_n(X)^2] \to 0$ by the condition (3).

Therefore, by applying Lemma 2.4.2, we have

$$\lim_{n\to\infty} D_{\mathrm{KL}}\left(\mathbb{P}_I \parallel \mathbb{Q}_I^{(n)}\right) = 0 \quad \text{if and only if} \quad \varphi(I) = -\phi'(y^*).$$

Since $0 < -\phi'(y^*) < \infty$, $\mathbb{P}_I$ satisfies the definition of the canonical probability distributions in (2.31). □

2.5 Applications

2.5.1 Gibbs measure on the phase space

Definition 2.5.1. *Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\mathbf{V} = (V_1, V_2, ..., V_n) : \Omega \to \mathbb{R}^n$ be a measurable function and let π_1, π_2, and π be three projection maps defined on $\mathbb{R}^n$ such that*

$$\pi_1(\mathbf{V}) = \mathbf{U} = (V_1, V_2, ..., V_k), \quad \pi_2(\mathbf{V}) = \mathbf{W} = (V_{k+1}, V_{k+2}..., V_n), \quad \pi(\mathbf{V}) = \mathbf{V}. \tag{2.142}$$

Assume there exist measurable functions $e_1 : \mathbb{R}^k \to \mathbb{R}^+$, $e_2 : \mathbb{R}^{n-k} \to \mathbb{R}^+$, *and* $e : \mathbb{R}^n \to \mathbb{R}^+$ *such that*

$$(e_1 \circ \pi_1)(\mathbf{V}) = e_1(\mathbf{U}), \quad (e_2 \circ \pi_2)(\mathbf{V}) = e_2(\mathbf{W}), \quad (e \circ \pi)(\mathbf{V}) = e(\mathbf{V}).$$

Therefore, random variables and induced measures can be defined through the following maps:

$$(\Omega, \mathcal{F}, \mathbb{P}) \xrightarrow{\mathbf{V}} (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu) \xrightarrow{\pi_1} (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \nu_1) \xrightarrow{e_1} (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+), \lambda_1)$$
$$(\Omega, \mathcal{F}, \mathbb{P}) \xrightarrow{\mathbf{V}} (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu) \xrightarrow{\pi_2} (\mathbb{R}^{n-k}, \mathcal{B}(\mathbb{R}^{n-k}), \nu_2) \xrightarrow{e_2} (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+), \lambda_2).$$

Definition 2.5.2. *Let* $e_1 \circ \pi_1, e_2 \circ \pi_2$, *and* $e \circ \pi$ *be the functions given in Definition 2.5.1. Define* $e_1 \circ \pi_1$ *and* $e_2 \circ \pi_2$ *to be* additive *on* $\mathbf{V}$ *if*

$$e_1 \circ \pi_1(\mathbf{V}) + \ e_2 \circ \pi_2(\mathbf{V}) = e \circ \pi(\mathbf{V}). \tag{2.143}$$

Theorem 2.5.1. *Suppose* $e_1 \circ \pi_1$ *and* $e_2 \circ \pi_2$ *are additive on* $\mathbf{V}$ *and suppose* $e_1(\mathbf{U}), e_2(\mathbf{W})$ *are continuous nonnegative independent random variables. Denote* $X := e_1(\mathbf{U}),\ Y := e_2(\mathbf{W}),\ Z := e(\mathbf{V})$ *and let* $I = [h, h+\delta]$ *be a finite interval. Assume the following conditions hold:*

1. $\mathbb{E}[X^2] = \epsilon_n^2$, *where* $\epsilon_n \to 0$.

2. *For all* $y \in \mathbb{R}^+$, *there exists a nonnegative integrable function* $\Gamma \in C^2(\mathbb{R}^+)$ *such that*

$$P\left(Y \in [y, y+\delta]\right) = \frac{\displaystyle\int_y^{y+\delta} \Gamma(s) ds}{\displaystyle\int_{\mathbb{R}^+} \Gamma(s) ds} \quad \text{and} \quad \left| \frac{\partial^2 \log P\big(Y \in [y, y+\delta]\big)}{\partial y^2} \right| < \infty. \tag{2.144}$$

3. $I \subset supp(\Gamma)$ *and* $\Gamma'(y) \geq 0$, *for* $y \in I$.

Then we have

$$\sup_{S \in \mathcal{B}(\mathbb{R}^+)} \left| \mathbb{P}\left(e_1(\mathbf{U}) \in S \mid Z \in I\right) - \int_{e_1(\mathbf{u}) \in S} A e^{-\psi(I) e_1(\mathbf{u})} \nu_1(d\mathbf{u}) \right| = O(\epsilon_n), \tag{2.145}$$

where $\psi(I) = \left. \dfrac{\partial \log \displaystyle\int_y^{y+\delta} \Gamma(s) ds}{\partial y} \right|_{y=h}$.

Proof. Since the functions $e_1 \circ \pi_1, e_2 \circ \pi_2$ are *additive* on $\mathbf{V}$, we have

$$X + Y = e_1(\mathbf{U}) + e_2(\mathbf{W}) = (e_1 \circ \pi_1)(\mathbf{V}) + (e_2 \circ \pi_2)(\mathbf{V}) = (e \circ \pi)(\mathbf{V}) = e(\mathbf{V}) = Z.$$

Since $X + Y = Z$ and they are corresponding to X_n, Y_n, Z_n in Theorem 2.3.1, respectively, it suffices to show that all the conditions in Theorem 2.3.1 are satisfied for X, Y, and Z.

1. For all $y \in \mathbb{R}^+$, since $\Gamma(y) \in C^2(\mathbb{R}^+)$, $\left|\dfrac{\partial^2 P\left(Y \in [y, y+\delta]\right)}{\partial^2 y}\right|$ exists and is bounded on $\mathbb{R}^+$. And $\left|\dfrac{\partial^2 \log P\left(Y \in [y, y+\delta]\right)}{\partial^2 y}\right|$ is bounded on $\mathbb{R}^+$ by (2.144). Therefore, (2.36) holds.

2. Since $I \subset supp(\Gamma)$, there exists $\delta_1 > 0$ such that $P\left(Y \in I\right) \geq \delta_1$. And we can derive

$$\left.\frac{\partial \log P(Y \in [y, y+\delta])}{\partial y}\right|_{y=h} = \frac{\Gamma(h+\delta) - \Gamma(h)}{\displaystyle\int_h^{h+\delta} \Gamma(s) ds}. \tag{2.146}$$

   Again, since $I \subset supp(\Gamma)$, and the nonnegative function $\Gamma(y) \in C^2(\mathbb{R}+)$, $\Gamma'(y) \geq 0$, for $y \in I$, we can check that there exists a positive constant C such that

$$0 \leq \frac{\partial \log P\big(Y \in [y, y+\delta]\big)}{\partial y} \leq C \quad \text{for } [y, y+\delta] \subset I, \tag{2.147}$$

   hence (2.37) holds for $D = I$. Furthermore, since X and Y are independent, $b_n = 0$. Therefore, (2.38) holds.

3. Since X and Y are supported on $\mathbb{R}^+$, there exists $\delta_2 > 0$ such that

$$P\left(Z \in [z, z+\delta]\right) \geq \delta_2 \quad \text{for } [y, y+\delta] \subset \mathbb{R}^+,$$

   hence (2.39) holds.

Therefore, all of the conditions hold for $D = I$ in Theorem 2.3.1, we can apply it with $a_n = \epsilon_n^2,\ b_n = 0$, and Pinsker's inequality (2.22) to get

$$\sup_{S \in \mathcal{B}(\mathbb{R}^+)} \left| \mathbb{P}\left(X \in S \mid Z \in I\right) - \int_{x \in S} A e^{-\psi(I)x} f_X(x) dx \right| = O(\epsilon_n), \tag{2.148}$$

where $\psi(I) = \left.\dfrac{\partial \log \displaystyle\int_y^{y+\delta} \Gamma(s)ds}{\partial y}\right|_{y=h}$. Then applying a change of variables

$$\int_{x\in S} Ae^{-\psi(I)x} f_X(x)dx = \int_{x\in S} Ae^{-\psi(I)x}\lambda_1(dx) = \int_{e_1(\mathbf{u})\in S} Ae^{-\psi(I)e_1(\mathbf{u})}\nu_1(d\mathbf{u}) \tag{2.149}$$

to (2.148), we obtain Equation (2.145). It completes the proof. □

In statistical mechanics, the induced measure $\nu_1(d\mathbf{u})$ in phase space is often considered as the Lebesgue measure $d\mathbf{u}$ normalized by the total volume of the phase space Λ (Here we assume it is finite). Therefore, for the random vector $\mathbf{U}$, we have its density

$$\hat{A}e^{-\psi(I)e_1(\mathbf{u})} \quad \text{with respect to } d\mathbf{u}, \tag{2.150}$$

where $\hat{A} = A/\Lambda$ is the corresponding normalization factor.

The assumption $\nu_1(d\mathbf{u}) = d\mathbf{u}/\Lambda$ for the phase space has already implied that all microstates are *equally probable* when the system is unconstrained. It is a reasonable *prior* probability for $\mathbf{U}$ by a symmetry of a physical system when we do not have any previous information about it. For the random variable X (e.g. energy), its density $f_X(x)$ is referred to *prior* probability for X when it is unconstrained. Based on the principle of equal a priori probabilities of microstates in the phase space, we can show that $f_X(x) = \gamma(x)/\Lambda$, where $\gamma(x)$ is the Lebseque measure of the surface area of microstates when the energy is fixed on x (i.e. $e_1(\mathbf{U}) = x$). This can be verified by

$$\int_{x\in S} f_X(x)dx = \int_{e_1(\mathbf{u})\in S} \nu_1(d\mathbf{u}) = \frac{1}{\Lambda}\int_{e_1(\mathbf{u})\in S} d\mathbf{u} = \frac{1}{\Lambda}\int_{x\in S}\gamma(x)dx \quad \text{for all } S \in \mathcal{B}(\mathbb{R}^+). \tag{2.151}$$

Note that $\gamma(x)$ is also known as the *structure function of* X. In Theorem 2.5.1, we also make the same assumption for Y: $f_Y(y) = \Gamma(y)/\Lambda$, where $\Gamma(y)$ is the *structure function of* Y.

Therefore, the density of X can be written as

$$\hat{A}e^{-\psi(I)x}\gamma(x) \quad \text{with respect to } dx, \tag{2.152}$$

which can be interpreted as a uniform prior biased by an exponential weight $e^{-\psi(I)x}$ when the system is conditioned on some extra information. Note that Equation (2.150) is known as the density of Gibbs measure on the phase space and Equation (2.152) is known as the density of Gibbs measure on the energy of the system [78].

In the work of A. Ya. Khinchin [108], he assumed *conjugate distribution laws* for all systems. It is said that

$$f_X(x) = \frac{e^{-\alpha x}\gamma(x)}{\int e^{-\alpha s}\gamma(s)ds} \quad \text{and} \quad f_Y(y) = \frac{e^{-\alpha y}\Gamma(y)}{\int e^{-\alpha s}\Gamma(s)ds} \tag{2.153}$$

for some constant α. Those priors are more general than the uniform prior and they have some nice properties, e.g., for a proper α, it may guarantee integrability of $e^{-\alpha s}\gamma(s)$ when $\gamma(s)$ itself is not integrable. However, we can show that the choice of $e^{-\alpha x}$ term does not have influence on our results. Here is the proof sketch: Suppose $\delta = o(1)$,

$$\begin{aligned}\hat{\psi}(I) := \left.\frac{\partial \log P(Y \in [y, y+\delta])}{\partial y}\right|_{y=h} &= \left.\frac{\partial \log \displaystyle\int_y^{y+\delta} \Gamma(s)e^{-\alpha s}ds}{\partial y}\right|_{y=h} \\ &\approx \left.\frac{\partial \log \displaystyle\int_y^{y+\delta} \Gamma(s)ds}{\partial y}\right|_{y=h} - \alpha = \psi(I) - \alpha. \end{aligned} \tag{2.154}$$

By (2.153) and (2.154), we have

$$Af_X(x)e^{-\hat{\psi}(I)x} = \hat{A}\gamma(x)e^{-\alpha x}e^{-(\psi(I)-\alpha)x} = \hat{A}\gamma(x)e^{-\psi(I)x}. \tag{2.155}$$

Therefore, to choose priors as the structure functions multiplied by the exponential functions $e^{-\alpha x}$ for integrability is irrelevant to Gibbs measure. Indeed, it is the extra information (condition) giving rise to the exponential weight in Gibbs measure and the parameter of the exponential function is determined by

$$\psi(I) = \left.\frac{\partial \log \displaystyle\int_y^{y+\delta} \Gamma(s)ds}{\partial y}\right|_{y=h},$$

in which $\int_y^{y+\delta} \Gamma(s)ds$ represents the volume of microstates in the shell between y and $y+\delta$. The logarithm of it is known as the entropy of Y, denoted by $S_Y(y)$, hence we have

$$\psi(I) = \left.\frac{\partial S_Y(y)}{\partial y}\right|_{y=h}. \tag{2.156}$$

By Equation (2.156), we can identify $\frac{1}{\psi(I)}$ as the temperature defined in statistical mechanics [92, 120].

Corollary 2.5.2. *We can extend Theorem 2.5.1 to the model that the subsystem and its heat bath have strong interaction defined by non-additivity of energy functions in statistical mechanics. Assume there exists a measurable function $e_3 : \mathbb{R}^n \to \mathbb{R}^+$ such that*

$$(e_3 \circ \pi)(\mathbf{V}) = e_3(\mathbf{V}),$$

which means that this energy function e_3 could depend on the whole vector $\mathbf{V} = (V_1, V_2, ..., V_n)$ in the phase space. And suppose

$$e_1 \circ \pi_1(\mathbf{V}) + e_2 \circ \pi_2(\mathbf{V}) + e_3 \circ \pi(\mathbf{V}) = e \circ \pi(\mathbf{V}),$$

in which the existence of the extra term $e_3 \circ \pi(\mathbf{V})$ means that $e_1 \circ \pi_1$ and $e_2 \circ \pi_2$ are not additive *on $\mathbf{V}$ by Definition 2.5.1. Denote that $R := e_3(\mathbf{V})$. Recall that $\mathbf{V} = (\mathbf{U}, \mathbf{W})$ and $X = e_1(\mathbf{U}) = (e_1 \circ \pi_1)(\mathbf{V}),\ Y = e_2(\mathbf{W}) = (e_1 \circ \pi_2)(\mathbf{V}),\ Z = e(\mathbf{V}) = (e \circ \pi)(\mathbf{V})$. Then we have*

$$X + Y + R = Z. \tag{2.157}$$

In statistical mechanics, R is known as the interaction energy *caused by interaction between the subsystem and its heat bath. Based on this setup, we can define a new random variable $\hat{Y} := Y + R$, but $X, \hat{Y}$ are no longer independent since the random variable R may depend on both $\mathbf{U}, \mathbf{W}$ in the phase space. If we modify the condition* (2.144) *in Theorem 2.5.1 to guarantee the existence and boundedness of*

$$\left|\partial^{(k)} P\big(\hat{Y} \in [y, y+\delta] \mid X = x\big)\right| \text{ and } \left|\partial^{(k)} \log P\big(\hat{Y} \in [y, y+\delta] \mid X = x\big)\right|, \text{ for } k = 0, 1, 2, \tag{2.158}$$

in which the partial derivatives are with respect to both x and y, then we are able to apply Corollary 2.3.4 to this model. As the result (2.46) *in Corollary 2.3.4, this model with strong interaction will give rise to a new parameter $\phi(I)$ of the exponential weight which involves two terms: one is from fluctuations of the energy of the "new" heat bath $\hat{Y}$ (it combines the energy of the heat bath Y without interaction and the interaction energy R); and the other one is from the correlation of X and $\hat{Y}$.*

2.5.2 Integer-valued random variables and conditional Poisson distributions

In the following Theorem 2.5.3, we will show a limiting behavior of a sequence of conditional probabilities for a nonnegative integer-valued random variables K, which is conditioned on $K+\tilde{L}_n$, $\tilde{L}_n$ is a sequence of sums of i.i.d random variables ξ_i. This sequence of conditional probabilities has the same limiting behavior as its unconditional probability $P(K=k)$ weighted by an exponential factor. The most important result of this theorem is that the parameter of this exponential factor determined by a normal distribution rather than the distribution of ξ_i. By this result, we provide a very simple formula with an approximation error to approximate an intractable problem in calculating the conditional probability of an integer-valued random variable. And we give an example 2.5.1 to show an approximation formula for calculating the conditional probability of a Poisson random variable conditioned on the sum of that Poisson random variable with another Poisson random variable.

Theorem 2.5.3. *Let K be a nonnegative integer-valued random variable with $\mathbb{E}[K]<\infty$. Let $\tilde{L}_n=\sum_{i=1}^{n}\xi_i$, where $\{\xi_i\}_{i=1}^{n}$ are nonnegative i.i.d. random variables. K and $\tilde{L}_n$ are independent and denote $\tilde{H}_n := K+\tilde{L}_n$. Let $\mu=\mathbb{E}[\xi_i]$, $\sigma^2=\mathrm{Var}(\xi_i)$ and assume $\mathbb{E}[(\xi_i-\mu)^3]<\infty$. And let*

$$B_n=\sum_{k=0}^{\infty}\frac{1}{P(K=k)\exp\left(\frac{-\psi(I)k}{\sqrt{n}}\right)},\quad \psi(I)=\left.\frac{\partial\log P\big(Y\in[y,y+\delta]\big)}{\partial y}\right|_{y=-h},\quad \textit{and}\, Y\sim N(0,\sigma^2).$$

For every fixed finite interval $I=[-h,-h+\delta],\ h,\delta\in\mathbb{R}^+$, $-h+\delta\le 0$, and $2\delta/\sigma^2<\psi(I)$,

$$\sup_k\left|P(K=k\mid \tilde{H}_n\in n\mu+\sqrt{n}I)-B_nP(K=k)\exp\left(\frac{-\psi(I)k}{\sqrt{n}}\right)\right|=O(\frac{1}{\sqrt{n}}). \tag{2.159}$$

Proof. Let $K_n:=\frac{K}{\sqrt{n}}$, $L_n:=\frac{\tilde{L}_n-n\mu}{\sqrt{n}}$ and $H_n:=\frac{\tilde{H}_n-n\mu}{\sqrt{n}}$. We have $K_n+L_n=H_n$. By the Central Limit Theorem, L_n converges in distribution to Y. Furthermore, since $(\xi_i-\mu)$ has finite second and third moments, by Berry-Esseen Theorem 2.2.4,

$$\sup_k\left|P_{L_n}\left(I-\frac{k}{\sqrt{n}}\right)-P_Y\left(I-\frac{k}{\sqrt{n}}\right)\right|=O\left(\frac{1}{\sqrt{n}}\right). \tag{2.160}$$

Since $\mathbb{E}\left[K_n\right]\to 0$, we have K_n converges to 0 in probability. By Slutsky's Theorem 2.2.5, H_n converges to Y in distribution. By Corollary 2.2.6, we can also get

$$P_{H_n}(I) = P_Y(I) + O\left(\frac{1}{\sqrt{n}}\right). \tag{2.161}$$

By (2.160) and (2.161),

$$\begin{aligned} P_{K|\tilde{H}_n}\left(k; n\mu + \sqrt{n}I\right) = P_{K|H_n}(k; I) = P_K(k)\frac{P_{L_n}\left(I - \frac{k}{\sqrt{n}}\right)}{P_{H_n}(I)} = P_K(k)\frac{P_Y\left(I - \frac{k}{\sqrt{n}}\right) + O\left(\frac{1}{\sqrt{n}}\right)}{P_Y(I) + O\left(\frac{1}{\sqrt{n}}\right)} \\ = P_K(k)\frac{P_Y\left(I - \frac{k}{\sqrt{n}}\right)}{P_Y(I)} + O\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \tag{2.162}$$

in which we use the fact $Y \sim N(0, \sigma^2)$ and $P(-h \leq Y \leq -h + \delta)$ is bounded from below. Moreover, since $P_K(k) \leq 1$, the term $O\left(\frac{1}{\sqrt{n}}\right)$ in (2.162) is independent of k. Let $\tilde{Y}_n \sim N(n\mu, n\sigma^2)$ and $\tilde{Z}_n := K + \tilde{Y}_n$. Then we have

$$K_n + Y_n = Z_n, \quad \text{where} \quad Y_n := \frac{\tilde{Y}_n - n\mu}{\sqrt{n}} \text{ and } Z_n := \frac{\tilde{Z}_n - n\mu}{\sqrt{n}}. \tag{2.163}$$

Note that $Y_n = Y \sim N(0, \sigma^2)$ and Z_n converges in distribution to Y. Similar to (2.162),

$$P_{K|\tilde{Z}_n}\left(k; n\mu + \sqrt{n}I\right) = P_K(k)\frac{P_Y\left(I - \frac{k}{\sqrt{n}}\right)}{P_Y(I)} + O\left(\frac{1}{\sqrt{n}}\right). \tag{2.164}$$

Applying the triangle inequality to (2.162) and (2.164), we finally obtain

$$\sup_k \left| P_{K|\tilde{H}_n}\left(k; n\mu + \sqrt{n}I\right) - P_{K|\tilde{Z}_n}\left(k; n\mu + \sqrt{n}I\right) \right| = O\left(\frac{1}{\sqrt{n}}\right). \tag{2.165}$$

Now, it remains to show that the convergence rate of

$$\sup_k \left| P_{K|\tilde{Z}_n}\left(k; n\mu + \sqrt{n}I\right) - B_n P_K(k) \exp\left(\frac{-\psi(I)k}{\sqrt{n}}\right) \right|. \tag{2.166}$$

Then it suffices to show that all the conditions in Theorem 2.3.1 are satisfied for K_n, Y_n, Z_n, then we can apply Theorem 2.3.6.

First, we can check that $\mathbb{E}[K_n^2] = a_n,\ a_n = o(1)$:

$$\mathbb{E}[K_n^2] = \frac{1}{n}\mathbb{E}[K^2] = O\left(\frac{1}{n}\right). \tag{2.167}$$

Second, by change of variables,

$$P_{K|\tilde{H}_n}\left(k; n\mu + \sqrt{n}I\right) = P_{K_n|H_n}\left(\frac{k}{\sqrt{n}}; I\right). \tag{2.168}$$

And we can define the set S in terms of the value for K as below:

$$S = \{k : k \in \mathbb{N},\ \mathbb{P}(K = k) > 0\}$$

such that for all $k \in S$, $P(K_n = \frac{k}{\sqrt{n}}) > 0$. Choose $d > 0$ such that $I = [-h, -h+\delta] \subseteq D = (-d, 0)$. Below we follow every steps in Theorem 2.3.1 with slight modifications:

1. For all $y \in \mathbb{R}$, $Y_n = Y \sim N(0, \sigma^2)$, by the formula of the density of normal distribution, we have

$$\frac{\partial^2 P\big(Y \in [y, y+\delta]\big)}{\partial y^2} = f'_Y(y+\delta) - f'_Y(y) \tag{2.169}$$

and

$$\frac{\partial^2 \log P\big(Y \in [y, y+\delta]\big)}{\partial y^2} = \frac{f'_Y(y+\delta) - f'_Y(y)}{P\big(Y \in [y, y+\delta]\big)} - \left(\frac{f_Y(y+\delta) - f_Y(y)}{P\big(Y \in [y, y+\delta]\big)}\right)^2, \tag{2.170}$$

so we can check (2.169) exist and are uniformly bounded. For (2.170), we modify the boundedness slightly and the details of proof are provided in Appendix 2.6.3. Therefore, (2.36) with a slight modification holds.

2. Since $Y_n = Y \sim N(0, \sigma^2)$, there exist positive constants δ_1 and C depending on y such that $P\big(Y \in [y, y+\delta]\big) \geq \delta_1$ and $0 \leq \dfrac{\partial \log P\big(Y \in [y, y+\delta]\big)}{\partial y} \leq C$ for every $[y, y+\delta] \subset D$. Therefore (2.37) holds. Since K_n and Y_n are independent, we have $b_n = 0$. Therefore (2.38) holds.

3. Since $Z_n \to Y$ in distribution where $Y \sim N(0, \sigma^2)$, there exists $\epsilon_n(z) \to 0$ such that

$$P\big(Z_n \in [z, z+\delta]\big) = P\big(Y \in [z, z+\delta]\big) + \epsilon_n(z).$$

Since $P\big(Y \in [z, z+\delta]\big)$ is bounded from below for $[z, z+\delta] \subset D$, there exists a positive constant $\delta_2(z)$ such that $P\big(Z_n \in [z, z+\delta]\big) \geq \delta_2 > 0$ for all $[z, z+\delta] \subset D$. Then the second inequality in (2.39) holds.

To apply Theorem 2.3.6, we then obtain

$$\sup_{k\in S}\left|P_{K_n|Z_n}\left(\frac{k}{\sqrt{n}};I\right)-B_nP_{K_n}\left(\frac{k}{\sqrt{n}}\right)\exp\left(-\psi(I)\frac{k}{\sqrt{n}}\right)\right|=O(\frac{1}{n}), \tag{2.171}$$

where

$$\psi(I)=\frac{\partial\log P_Y\big([y,y+\delta]\big)}{\partial y}\bigg|_{y=-h}\quad\text{and}\quad Y\sim N(0,\sigma^2). \tag{2.172}$$

By change of variable, we then obtain

$$\sup_{k}\left|P_{K|\tilde{Z}_n}\big(k;n\mu+\sqrt{n}I\big)-B_nP_K(k)\exp\left(\frac{-\psi(I)k}{\sqrt{n}}\right)\right|=O(\frac{1}{n}), \tag{2.173}$$

where

$$B_n=\frac{1}{\sum_{k\in S}P_{K_n}(k/\sqrt{n})\exp\left(-\psi(I)k/\sqrt{n}\right)}=\frac{1}{\sum_k P_K(k)\exp\left(-\psi(I)k/\sqrt{n}\right)}.$$

By applying triangle inequality to (2.165) and (2.173), we can obtain (2.159) in the theorem.

□

Finally we apply Theorem 2.5.3 to a concrete example.

Example 2.5.1. *Let $\lambda,\mu>0$ be two constants. Consider two independent random variables $K\sim$ Pois(λ) and $\tilde{L}_n\sim$ Pois$(n\mu)$. Let $\tilde{H}_n:=K+\tilde{L}_n$. For every fixed finite interval I which follows from Theorem 2.5.3, we can show that*

$$\sup_{k}\left|P\left(K=k\mid\tilde{H}_n\in n\mu+\sqrt{n}I\right)-B_nP(K=k)\exp\left(\frac{-\psi(I)k}{\sqrt{n}}\right)\right|=O(\frac{1}{\sqrt{n}}),$$

where $B_n=\displaystyle\sum_{k=0}^{\infty}\frac{1}{P(K=k)\exp\left(\frac{-\psi(I)k}{\sqrt{n}}\right)}$ *and* $\psi(I)=\dfrac{\partial\log P\big(Y\in[y,y+\delta]\big)}{\partial y}\bigg|_{y=-h}$, $Y\sim N(0,\mu)$.

Proof. By the property of Poisson random variables, we can decompose $\tilde{L}_n$ as $\tilde{L}_n=\sum_{i=1}^n\xi_i$, where $\{\xi_i,1\leq n\}$ are independent Poisson random variables with mean μ and variance μ. We can check that all conditions are satisfies in Theorem 2.5.3. Hence Theorem 2.5.3 can be applied. □

2.5.3 Emergence of temperature (conditioned on the scale of large deviations)

In this section, we define the parameter $\frac{1}{\varphi(I)}$ in the exponential function $e^{-\varphi(I)x}$ as the *temperature* of the canonical distribution. Consider a sequence of conditional probabilities for a function of a subsystem represented by X in contact with its heat bath represented by $\tilde{Y_n} = \sum_{i=2}^{n} X_i$, where X_i are i.i.d. and X_i has the same distribution as X, and X_i, X are independent. Suppose that the total energy $\tilde{Z}_n = X + \tilde{Y}_n$ is conditioned on the scale of large deviations from its mean, we will show that the *temperature* $\frac{1}{\varphi(I)}$ is an emergent parameter uniquely determined by the rate function of $\frac{\tilde{Y}_n}{n}$.

Definition 2.5.3. *Let X be a nonnegative and nonconstant continuous random variable with $\mathbb{E}[X^4] < \infty$, and let $\tilde{Y}_n := \sum_{i=2}^{n} X_i$, where all random variables in $\{X_i\}_{i=2}^{n} \cup \{X\}$ are i.i.d.. Denote $\tilde{Z}_n := X + \tilde{Y}_n$. Consider an interval $I = [d, d+\delta],\ d \in \mathbb{R},\ \delta > 0$ with $\mathbb{E}[X] \notin I$, and a function $\varphi : I \to \mathbb{R}$ such that $0 < \varphi(I) < \infty$. Let $\mathbb{P}_I$ be a probability measure with density function $Af_X(x)e^{-\varphi(I)x}$, where*

$$\frac{1}{A} = \int_{\mathbb{R}^+} f_X(x)e^{-\varphi(I)x}dx.$$

Let $\mathbb{Q}_I^{(n)}$ be a sequence of probability measures with density functions $f_{X|\tilde{Z}_n}(x; nI)$.

Theorem 2.5.4. *Denote $Y_n := \frac{\tilde{Y}_n}{n}$, $X_n := \frac{X}{n}$, and $Z_n := X_n + Y_n$, and let $I - \frac{x}{n} = \{y - \frac{x}{n},\ y \in I\}$. Assume the following conditions hold:*

1. $\left|\dfrac{f_{X|Z_n}(x; I)}{f_X(x)}\right|$ *is uniformly bounded on $\mathbb{R}^+$.*
2. $|\log P_{Y_n}(I) - \log P_{Z_n}(I)|$ *converges to a finite constant as $n \to \infty$.*
3. *There exists a function $\phi(y) \in C^2(D)$, where D is an open interval containing I, with $-\infty < \phi'(y) < 0$, for $y \in I$, such that*

$$\log P_{Y_n}\left(I - \frac{x}{n}\right) = -n\phi\left(y^* - \frac{x}{n}\right) + s_n\left(I - \frac{x}{n}\right), \quad \text{for } I - \frac{x}{n} \subset D, \tag{2.174}$$

where $y^* = \{y : \inf_{y \in I} \phi(y)\}$, $\left|\dfrac{s_n(I - \frac{x}{n}) - s_n(I)}{s_n(I)}\right| = O\left(\dfrac{x}{n}\right)$, and $|s_n(I')| = o(n)$ for all $I' \subset D$.

Then

$$D_{\mathrm{KL}}\left(\mathbb{P}_I \parallel \mathbb{Q}_I^{(n)}\right) \to 0 \qquad \textit{if and only if} \qquad \varphi(I) = -\phi'(y^*), \quad y^* = \{y : \inf_{y\in I} \phi(y)\}. \tag{2.175}$$

Corollary 2.5.5. *The conditions* (1) - (3) *formulated in Theorem 2.5.4 are technical, so we would like to characterise and verbally describe the underlying meaning and interpretation of them: The condition* (1) *can be written as*

$$\left|\frac{f_{X|Z_n}(x;I)}{f_X(x)}\right| = \left|\frac{f_{X_n,Y_n}\left(\frac{x}{n}, I-\frac{x}{n}\right)}{f_{X_n}(\frac{x}{n})f_{Y_n}(I-\frac{x}{n})}\right| \quad \textit{is uniformly bounded on } \mathbb{R}^+,$$

in which the right hand side is related to the correlation of X_n and Y_n, therefore, this condition means that the interaction between X_n and Y_n is regulated; The condition (2) *is corresponding to the setup that X_n is small relative to Z_n (hence the distributions of Y_n and Z_n have the same asymptotic behavior), specifically, that finite constant can be chosen to be zero (we provide a more general condition in this theorem); The condition* (3) *means that Y_n converges to a constant satisfying the large deviation principle with the rate function ϕ and the remainder term s_n.*

Proof. The proof of Theorem 2.5.4 is just the application of Theorem 2.3.9, so we will show that all conditions in Theorem 2.3.9 are satisfied. First, Condition (1) in Theorem 2.3.9 follows from Condition (1), and $\mathbb{E}[X^4] < \infty$ is assumed in this theorem. Second, Condition (2) in Theorem 2.3.9 follows from (i) $Y_n \to \mathbb{E}[X]$ in probability by the law of large numbers, (ii) $\mathbb{E}[X] \notin I$ by Definition 2.5.3, and (iii) the Condition (3) in this theorem.

Third, since I is closed and contained in an open interval D, there exists a constant $d \in \mathbb{R}^+$ such that $I - \dfrac{x}{n} \subset D$ for $x \in [0, nd]$. Therefore, by Condition (3),

$$\log P_{Y_n}\left(I - \frac{x}{n}\right) = -n\phi\left(y^* - \frac{x}{n}\right) + s_n\left(I - \frac{x}{n}\right), \quad y^* = \left\{y : \inf_{y\in I} \phi(y)\right\}. \tag{2.176}$$

Since $\left[y^*, y^* - \frac{x}{n}\right] \subseteq D$ and $\phi \in C^2(D)$, by Taylor's expansion,

$$\phi\left(y^* - \frac{x}{n}\right) = \phi(y^*) - \phi'(y^*)\frac{x}{n} + O\left(\frac{x^2}{n^2}\right) \quad \text{for all } x \in [0, nd]. \tag{2.177}$$

By Condition (2) and (3), there exists a sequence $\epsilon_n \to 0$ and a constant k such that

$$\log P_{Z_n}(I) = \log P_{Y_n}(I) + k + \epsilon_n = -n\phi(y^*) + s_n(I) + k + \epsilon_n. \tag{2.178}$$

By Condition (3), we have

$$\left| s_n \left(I - \frac{x}{n}\right) - s_n(I) \right| = |s_n(I)|\, O\left(\frac{x}{n}\right) = O(\delta_n x), \tag{2.179}$$

in which $\delta_n \to 0$. By the results of (2.176), (2.177), (2.178), and (2.179), we obtain

$$\log\left(\frac{P_{Y_n}\left(I - \frac{x}{n}\right)}{P_{Z_n}(I)}\right) = \log\left(\frac{\exp\left[-n\phi\left(y^* - \frac{x}{n}\right)\right]}{\exp\left[-n\phi(y^*)\right]}\right) + O\left(\frac{x^2}{n}\right) + O(\delta_n x) + \epsilon_n \quad \text{on} \quad I_n = [0, nd]. \tag{2.180}$$

Let $r_n(x) := O\left(\frac{x^2}{n}\right) + O(\delta_n x) + \epsilon_n$, we can check that (i) $\left|r_n(x) e^{-\xi x}\right|$ uniformly bounded on $\mathbb{R}^+$ for any $\xi > 0$, and (ii) $\mathbb{E}\left[r_n(X)^2\right] \to 0$ since $\mathbb{E}\left[X^4\right] < \infty$ by Definition 2.5.3. Hence, $r_n(x), d_n, \phi$ satisfy Condition (3) in Theorem 2.3.9. Therefore, we have checked that all of the conditions in Theorem 2.3.9 hold, then we can apply it to get

$$D_{\mathrm{KL}}\left(\mathbb{P}_I \parallel \mathbb{Q}_I^{(n)}\right) \to 0 \quad \text{if and only if} \quad \varphi(I) = -\phi'(y^*). \tag{2.181}$$

$\square$

By Cramér's Theorem 2.2.3, the existence of the function $\phi(y)$ in Condition (3) is from the existence of the rate function of $Y_n = \sum_{i=1}^{n-1} X_i / n$. Let set $D_\phi := \{y \in \mathbb{R} : \phi(y) < \infty\}$ and we can choose $D = \operatorname{int}(D_\phi)$. By the properties of rate functions in Appendix 2.6.1, we have

$$\phi(y) \in C^2(D) \quad , \quad \phi(y) \text{ is convex on } D, \tag{2.182}$$

and $-\infty < \phi'(y) < 0$ for $y \in I \subset D$ if the interval I is chosen on the left side of the mean of Y_n. By Cramér's Theorem, the rate function satisfies

$$\log P_{Y_n}\left(I - \frac{x}{n}\right) = -n\phi\left(y^* - \frac{x}{n}\right) + o(n), \quad \text{for } I - \frac{x}{n} \subset D. \tag{2.183}$$

Comparing (2.183) with Condition (3), Theorem 2.5.4 requires an explicit form of the remainder:

$$\log P_{Y_n}\left(I - \frac{x}{n}\right) = -n\phi\left(y^* - \frac{x}{n}\right) + s_n\left(I - \frac{x}{n}\right), \quad \text{for } I - \frac{x}{n} \subset D, \tag{2.184}$$

where $\left|\frac{s_n(I-\frac{x}{n})-s_n(I)}{s_n(I)}\right| = O\left(\frac{x}{n}\right)$, and $|s_n(I')| = o(n)$ for all $I' \subset D$. This stronger condition guarantees the "if and only" if statement (2.175).

The following is our discussion on the connection between Theorem 2.5.4 and Van Campenhout and Cover's Theorem 2.2.2. In Theorem 2.5.4, if the condition is on the scale of large deviations, then the conditional density

$$f_{X|\tilde{Z}_n}(x; nI),\ n\mu \notin nI$$

can be approximated by the (normalized) product of its unconditional density $f_X(x)$ and an exponential function $e^{-\lambda x}$. This parameter $\lambda = \phi'(y^*)$ is unique and determined by the first derivative of the rate function evaluated at $y^* = \inf_{y \in I} \phi(y)$. It implies that we are able to find λ directly from the rate function without using the maximum entropy principle. Furthermore, by the pair of reciprocal equations (2.17):

$$\phi'(y^*) = \lambda \quad \text{if and only if} \quad A'(\lambda) = y^*, \tag{2.185}$$

which means the parameter λ we find by the derivative of the rate function (left side of (2.185)) is also the solution of the derivative of the free energy function A under the constraint $= y^*$ (right side of (2.185)).

Therefore, using the maximum entropy principle under the first moment constraint to find good approximations of conditional density (Van Campenhout and Cover's approach) is a natural consequence of the emergent behavior of

$$\log\left(\frac{f_{X|\tilde{Z}_n}(x; nI)}{f_X(x)}\right). \tag{2.186}$$

And this emergent behavior gives rise to a large deviation function that uniquely determines the parameter of the exponential weight. As we discussed in the Section 2.2, we apply the large deviation principle directly to the distribution of a the heat bath

$$Y_n = \frac{\tilde{Y}_n}{n} = \frac{1}{n}\sum_{i=2}^{n} X_i.$$

On the other hand, the Gibbs conditioning principle uses the large deviation principle for emprical measures

$$L_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}.$$

Denote that $X_1 := X$. Then the limit problem of the sequence of probability measures $\mathbb{Q}_I^{(n)}$ with density functions

$$f_{X|Z_n}(x;I), \quad \text{where } Z_n = X + Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

and the limit problem of the sequence of emprical measures

$$\mathbb{E}\left[L_n \mid L_n \in \Gamma\right], \quad \text{where } L_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i} \text{ and } \Gamma = \left\{\gamma : \int x\gamma(dx) \in I\right\}$$

are just two sides of the same coin. Eventually, they both give arise to a limit as a canonical distirbution with the density

$$f_X(x)e^{-\lambda x}.$$

In conclusion, our approach generates λ by the large deviation rate function of the heat bath Y_n and the Gibbs conditioning principle solves λ by minimizing the relative entropy which is the large deviation rate function of sampling. These two approaches are connected by the reciprocal equations (2.185) through the Legendre transform.

2.5.4 Emergence of temperature (conditioned on the scale of Gaussian fluctuations)

Similar to Section 2.5.3, in this section, we define the parameter $\frac{1}{\beta_n\psi(I)}$ in the exponential function $e^{-\beta_n\psi(I)x}$ as the *temperature* of the canonical distribution and consider a sequence of conditional probabilities for a function of a subsystem represented by X in contact with its heat bath represented by $\tilde{Y}_n = \sum_{i=2}^{n} X_i$, X_i are i.i.d. and X_i has a same distribution as X, and X, X_i are independent. In comparison with Section 2.5.3, here we suppose that the total energy $\tilde{Z}_n := X + \tilde{Y}_n$ is conditioned on the scale of Gaussian fluctuations. We will show that the *temperature* $\frac{1}{\beta_n\psi(I)}$ is an emergent parameter uniquely determined by a normal distribution $N(0,\sigma^2)$, where σ^2 is the variance of X.

Definition 2.5.4. *Let X be a nonnegative and nonconstant continuous random variable with $\mathbb{E}[X^4] < \infty$, and let $\mu = \mathbb{E}[X]$, σ^2 be the variance of X. Let $\tilde{Y}_n = \sum_{i=2}^n X_i$, where all random variables in $\{X_i\}_{i=2}^n \cup \{X\}$ are i.i.d.. Denote $\tilde{Z}_n := X + \tilde{Y}_n$. For an interval $I = [d, d+\delta],\ d \in \mathbb{R}, \delta > 0$ and a function $\psi : I \to \mathbb{R}$ such that $0 < \psi(I) < \infty$. Let $\mathbb{P}_I^{(n)}$ be a sequence of probability measures with density functions $A_n f_X(x) e^{-\frac{\psi(I)}{\sqrt{n}}x}$, where*

$$\frac{1}{A_n} = \int_{\mathbb{R}^+} f_X(x) e^{-\frac{\psi(I)}{\sqrt{n}}x} dx$$

and let $\mathbb{Q}_I^{(n)}$ be a sequence of probability measures with density functions $f_{X|\tilde{Z}_n}(x; n\mu + \sqrt{n}I)$.

Theorem 2.5.6. *Denote $Y_n = \frac{\tilde{Y}_n - (n-1)\mu}{\sqrt{n}}$, $X_n = \frac{X}{\sqrt{n}}$, $Z_n = X_n + Y_n$, and let $I - \frac{x}{\sqrt{n}} = \left\{ y - \frac{x}{\sqrt{n}},\ y \in I \right\}$. Assume the following conditions hold:*

1. $\left| \dfrac{f_{X|Z_n}(x; I)}{f_X(x)} \right|$ *is uniformly bounded on $\mathbb{R}^+$.*

2. $Y_n \to Y$ *in distribution and* $\left. \dfrac{\partial \log P(Y \in [y, y+\delta])}{\partial y} \right|_{y=d} > 0,\ Y \sim N(0, \sigma^2).$

3. *There exists a sequence of functions $g_n : \mathbb{R} \to \mathbb{R}$ with*

$$\left| g_n(x) e^{-\frac{\xi}{\sqrt{n}}x} \right| \text{ uniformly bounded on } \mathbb{R}^+, \text{ for any } \xi > 0, \quad \text{and} \quad \mathbb{E}\left[g_n(X)^2\right] \to 0$$

such that

$$\log\left(\frac{P\left(Y_n \in I - \frac{x}{\sqrt{n}}\right)}{P\left(Z_n \in I\right)} \right) = \log\left(\frac{P\left(Y \in I - \frac{x}{\sqrt{n}}\right)}{P\left(Y \in I\right)} \right) + \frac{g_n(x)}{\sqrt{n}} \text{ on } I_n, \tag{2.187}$$

in which $I_n = [0, d_n]$ with $d_n = O(\sqrt{n})$.

Then

$$n D_{\mathrm{KL}}\left(\mathbb{P}_I^{(n)} \parallel \mathbb{Q}_I^{(n)} \right) \to 0 \quad \text{if and only if} \quad \psi(I) = \left. \frac{\partial \log P(Y \in [y, y+\delta])}{\partial y} \right|_{y=d}. \tag{2.188}$$

Corollary 2.5.7. *As Remark 2.5.5, the conditions* (1) - (3) *formulated in Theorem 2.5.6 are technical, so we would like to characterise and verbally describe the underlying meaning and interpretation of them: As Theorem 2.5.4, the condition* (1) *means that the interaction between X_n and Y_n is regulated; The condition* (2) *follows from the central limit theorem and we need to choose the interval $I = [d, d+\delta] \subset \mathbb{R}^-$ such that the partial derivative term is positive; The condition* (3) *combines the setup $X_n \to 0$ in probability and $Y_n \to Y$ in distribution, furthermore, the remainder term has a special form $\frac{g_n(x)}{\sqrt{n}}$.*

The proof of Theorem 2.5.6 is just the application of Theorem 2.3.8. We can check that all of the conditions in Theorem 2.3.8 are satisfied. Here we want to further discuss the equation (2.187) in Condition (3):

As the proof for Theorem 2.5.3, by Corollary 2.2.6 of Berry-Esseen theorem and Slusky's theorem, we have

$$\log\left(\frac{P\left(Y_n \in I - \frac{x}{\sqrt{n}}\right)}{P\left(Z_n \in I\right)}\right) = \log\left(\frac{P\left(Y \in I - \frac{x}{\sqrt{n}}\right)}{P\left(Y \in I\right)}\right) + O\left(\frac{1}{\sqrt{n}}\right) \quad \text{on } I_n. \tag{2.189}$$

However, it only guarantees the convergence of $\mathbb{P}_I^{(n)}$ and $\mathbb{Q}_I^{(n)}$ in $\|\cdot\|_\infty$ by Theorem 2.5.3. Compare Equation (2.189) with Condition (2), Theorem 2.5.6 requires an explicit form of the remainder:

$$\log\left(\frac{P\left(Y_n \in I - \frac{x}{\sqrt{n}}\right)}{P\left(Z_n \in I\right)}\right) = \log\left(\frac{P\left(Y \in I - \frac{x}{\sqrt{n}}\right)}{P\left(Y \in I\right)}\right) + \frac{g_n(x)}{\sqrt{n}} \quad \text{on } I_n, \tag{2.190}$$

and $\mathbb{E}[g_n(X)^2] \to 0$. This explicit form of remainder guarantees the "if and only" if statement (2.188).

We now discuss the connection between Theorem 2.5.6 and Zabell's Theorem 2.2.1. If the condition is on the scale of Gaussian fluctuations, Theorem 2.2.1 only tells us that the sequence of conditional distributions $F_{X|\tilde{Z}_n}(x; n\mu + \sqrt{n}I)$ should converge to its unconditional distribution $F_X(x)$. By our theorem 2.5.6, we have an explicit formula for the canonical distribution to approximate the conditional distribution well:

$$F_{X|\tilde{Z}_n}(x; n\mu + \sqrt{n}I) \approx \int_{-\infty}^{x} A_n f_X(s) e^{-\frac{\psi(I)}{\sqrt{n}} s} dx,$$

for a sufficiently large n, and it converges to $F_X(x)$ as $n \to \infty$ which is consistent with Zabell's Theorem 2.2.1. In addition, the parameter $\frac{\psi(I)}{\sqrt{n}}$ of the canonical distribution is uniquely determined if we require that the approximation is "good" enough, i.e. the KL-divergence of the conditional distribution from the canonical distribution converges to zero in the rate $o\left(\frac{1}{n}\right)$.

2.5.5 *Mathematical definitions of the heat bath*

In Section 2.3, we provided two limit theorems of a sequence of conditional probabilities to derive a unique canonical distribution as an emergent phenomenon. In Theorem 2.3.8, the emergent parameter in the exponential weight is uniquely determined by the limiting distribution of the heat bath $Y_n \to Y$ (note that in Theorem 2.3.8, Y_n follows from the appropriate shifting and scaling of the original heat bath $\tilde{Y}_n$) evaluated on the interval $I = [h, h+\delta]$ such as

$$\psi(I) = \left.\frac{\partial \log P(Y \in [y, y+\delta])}{\partial y}\right|_h .$$

Similarly, in Theorem 2.3.9, the emergent parameter in the exponential weight is uniquely determined by the large-deviation rate function of the heat bath $Y_n \to \mu$ (note that in Theorem 2.3.9, Y_n follows from the appropriate shifting and scaling of the original heat bath $\tilde{Y}_n$) evaluated on the interval $I = [h, h+\delta]$ such as

$$\varphi(I) = -\phi(y^*),$$

where ϕ is the rate function of Y_n and $y^* = \{y : \inf_{y \in I} \phi(y)\}$.

If we choose an interval $I' \subset I$, the parameter in the exponential weight may depend on I' in both of the limit theorems. However, since I' is just a subinterval of I, we expect that a well-defined *heat bath* should give rise to an *invariant temperature* of the canonical distribution by giving a constant parameter in the exponential weight no matter what subinterval I' we choose for it. In this section, we discuss two cases that follow from Theorem 2.3.8 and Theorem 2.3.9, respectively. Given a finite interval I, we first define the *subinterval invariant property* of a sequence of conditional distributions, then we provide three equivalent properties: (1) the subinterval invariant property of a sequence of conditional distributions (2) the invariant temperature property of the canonical distribution (3) the heat-bath property. Based on the equivalence of these three properties, we truly define the concept of "heat bath" in the language of mathematics.

Recall that X, $\tilde{Y}_n$, and $\tilde{Z}_n := X + \tilde{Y}_n$, are random variables from the definitions in Section 2.3. By proper shifting and scaling, let $X_n := \beta_n X$, $Y_n := \beta_n \left(\tilde{Y}_n - \mu_n\right)$, and $Z_n := X_n + Y_n$, where μ_n, β_n are positive sequences and $\beta_n = o(1)$.

For a finite interval $I = [h, h+\delta],\ h \in \mathbb{R}$ and $\delta > 0$, let $\mathbb{Q}_I^{(n)}$ be a sequence of probability measures with density functions $f_{X|Z_n}\big(x; I\big)$. The sequence of conditional probability measures $\mathbb{Q}_I^{(n)}$ represents our setup for the canonical ensemble, which should have a "nice" property such that the limiting behaviors of $\mathbb{Q}_{I'}^{(n)}$ and $\mathbb{Q}_I^{(n)}$ are the same for all subintervals $I' \subset I$. Hence we define this "nice" property as follows:

Definition 2.5.5. *Note that $\delta(\cdot,\cdot)$ represents the total variation distance of two probability measures. For any given interval $I' \subset I$,*

$$\frac{\delta\left(\mathbb{Q}_{I'}^{(n)}, \mathbb{Q}_I^{(n)}\right)}{\alpha_n} \to 0, \tag{2.191}$$

in which we take $\alpha_n = \beta_n$ for Theorem 2.3.8, and $\alpha_n = 1$ for Theorem 2.3.9. Then we say that the sequence of conditional probability measures $\mathbb{Q}_I^{(n)}$ has the subinterval invariant property *on the interval I.*

We start with our first theorem which follows Theorem 2.3.8. Recall that in Theorem 2.3.8, Y is a random variable such that $Y_n \to Y$ in distribution.

Theorem 2.5.8. *For a given interval $I' = [h', h'+\delta'],\ h' \in \mathbb{R}$, $\delta' > 0$, and $I' \subset I$, and a function $\psi : I' \to \mathbb{R}$, let $\tilde{\mathbb{P}}_{I'}^{(n)}$ be a sequence of probability measures with density functions*

$$\frac{f_X(x) e^{-\beta_n \psi(I') x}}{\displaystyle\int_{\mathbb{R}^+} f_X(x) e^{-\beta_n \psi(I') x} dx}, \tag{2.192}$$

where

$$\psi(I') = \left.\frac{\partial \log P(Y \in [y, y+\delta'])}{\partial y}\right|_{h'}. \tag{2.193}$$

Assume all of the conditions in Theorem 2.3.8 hold, then the following three statement are equivalent:

1. $\mathbb{Q}_I^{(n)}$ *has the subinterval invariant property on the interval* I.

2. $\tilde{\mathbb{P}}_{I'}^{(n)}$ *has a unique parameter (the invariant temperature property) such as*

$$\psi(I') = \psi(I) \quad \textit{for all} \quad I' \subset I.$$

3. $Y_n \to Y$ *in distribution and* Y *is a random variable with a distribution function*

$$P(Y \in [h', h' + \delta']) = \alpha(\delta') e^{\psi(I) h'} \quad \textit{for all } [h', h' + \delta'] \subset I, \tag{2.194}$$

where $\alpha : \mathbb{R}^+ \to \mathbb{R}$ *is a function.*

Proof. Since all of the conditions in Theorem 2.3.8 hold for all intervals $I' \subset I$ with (2.193), we can obtain that

$$\lim_{n\to\infty} \frac{D_{\mathrm{KL}}\left(\tilde{\mathbb{P}}_{I'}^{(n)} \parallel \mathbb{Q}_{I'}^{(n)}\right)}{\beta_n^2} = 0, \quad \text{for all } I' \subset I. \tag{2.195}$$

To prove ((1) ⇒ (2) ⇒ (3)): Assume the invariant temperature property holds, by applying the triangle inequality and Pinsker's inequality to Equation (2.195) and the assumption of the subinterval invariant property (2.191) with $\alpha_n = \beta_n$, we have that

$$\frac{\delta\left(\tilde{\mathbb{P}}_{I'}^{(n)}, \tilde{\mathbb{P}}_{I}^{(n)}\right)}{\beta_n} \to 0, \quad \text{for all } I' \subset I. \tag{2.196}$$

Following every step from (2.131) to (2.139) in the proof 2.4.2 for Lemma 2.4.3, we can get

$$\psi(I') = \psi(I), \quad \text{for all } I' \subset I. \tag{2.197}$$

By (2.193) and (2.197), we have

$$\left.\frac{\partial \log P(Y \in [y, y + \delta'])}{\partial y}\right|_{h'} \equiv \psi(I), \quad \text{for all } [h', h' + \delta'] \subset I, \tag{2.198}$$

which implies Y has a distribution

$$P(Y \in [h', h'+\delta']) = \alpha(\delta') e^{\psi(I)h'}, \quad \text{for all } [h', h'+\delta'] \subset I,$$

with some function $\alpha : \mathbb{R}^+ \to \mathbb{R}$.

To prove ((3) $\Rightarrow$ (2) $\Rightarrow$ (1)): By the assumption (3) that

$$P(Y \in [h', h'+\delta']) = \alpha(\delta') e^{\psi(I)h'}, \quad \text{for all } [h', h'+\delta'] \subset I,$$

with some function $\alpha : \mathbb{R}^+ \to \mathbb{R}$, and the equation (2.193), we can obtain that

$$\psi(I') = \psi(I), \quad \text{for all} \quad I' \subset I, \tag{2.199}$$

therefore, it implies

$$\frac{\delta\left(\tilde{\mathbb{P}}_{I'}^{(n)}, \tilde{\mathbb{P}}_{I}^{(n)}\right)}{\beta_n} = 0, \quad \text{for all } I' \subset I. \tag{2.200}$$

By applying the triangle inequality and Pinsker's inequality to (2.200) and (2.195), we have

$$\frac{\delta\left(\mathbb{Q}_{I'}^{(n)}, \mathbb{Q}_{I}^{(n)}\right)}{\beta_n} \to 0, \quad \text{for all } I' \subset I. \tag{2.201}$$

□

Next, we continue our analysis based on Theorem 2.3.9. Recall that in Theorem 2.3.9, $Y_n \to \mu$, for some constant μ, in probability and the sequence of laws of Y_n satisfies a large deviation principle with speed $1/\beta_n$ and rate function ϕ. The rate function $\phi \in C^2(D)$, where D is an open interval containing I, and

$$-\infty < \phi'(y) < 0, \text{ for all } y \in I. \tag{2.202}$$

Theorem 2.5.9. *For a given interval $I' = [h', h'+\delta']$, $h', \delta' \in \mathbb{R}$, $\delta' > 0$, and $I' \subset I$, and a function $\varphi : I' \to \mathbb{R}$, let $\mathbb{P}_{I'}$ be a probability measure with density function*

$$\frac{f_X(x) e^{-\varphi(I')x}}{\displaystyle\int_{\mathbb{R}^+} f_X(x) e^{-\varphi(I')x} dx}, \tag{2.203}$$

where

$$\varphi(I') = -\phi'(\hat{y}^*), \quad \hat{y}^* = \{y : \inf_{y \in I'} \phi(y)\}. \tag{2.204}$$

Assume all of the conditions in Theorem 2.3.9 hold, then the following three statements are equivalent:

1. $\mathbb{Q}_I^{(n)}$ *has subinterval invariant property on the interval* I.

2. $\mathbb{P}_{I'}$ *has a unique parameter (invariant temperature property) such as*

$$\varphi(I') = \varphi(I) \quad \textit{for all} \quad I' \subset I.$$

3. *Let* ϕ *be the large deviation rate function of* Y_n. ϕ *is a linear function such as*

$$\phi(y) = \phi'(y^*)y + c, \quad \textit{for all } y \in I, \tag{2.205}$$

where $y^* = \{y : \inf_{y \in I} \phi(y)\}$ *and* c *is some constant.*

Proof. Since all of the conditions in Theorem 2.3.9 hold for all intervals $I' \subset I$ with (2.204), we can obtain that

$$\lim_{n\to\infty} D_{\text{KL}}\left(\mathbb{P}_{I'} \parallel \mathbb{Q}_{I'}^{(n)}\right) = 0, \quad \text{for all } I' \subset I. \tag{2.206}$$

We first show $((1) \Rightarrow (2) \Rightarrow (3))$. The proof of $((1) \Rightarrow (2))$ follows from the proof of $((1) \Rightarrow (2))$ in Theorem 2.5.8, then we can get

$$\varphi(I') = \varphi(I), \quad \text{for all} \quad I' \subset I. \tag{2.207}$$

By (2.204) and (2.207),

$$\phi'(\hat{y}^*) = \phi'(y^*), \quad \text{for all } \hat{y}^* = \{y : \inf_{y \in I'} \phi(y)\} \text{ with } I' \subset I.$$

With the assumption (2.202): $-\infty < \phi'(y) < 0$, for all $y \in I$, and the properties of the rate function ϕ in Appendix 2.6.1, we have that

$$\phi'(y) \equiv \phi'(y^*), \quad \text{for all } y \in I,$$

which implies

$$\phi(y) = \phi'(y^*)y + c, \quad \text{for all } y \in I, \tag{2.208}$$

where c is some constant.

Next we prove ((3) $\Rightarrow$ (2) $\Rightarrow$ (1)). Equation (2.205) implies

$$\phi'(y) \equiv \phi'(y^*), \quad \text{for all } y \in I,$$

then we can obtain

$$\phi'(\hat{y}^*) = \phi'(y^*), \quad \text{for all } \hat{y}^* = \{y : \inf_{y \in I'} \phi(y)\} \text{ with } I' \subset I.$$

With (2.204), it implies

$$\varphi(I') = \varphi(I) \quad \text{for all } I' \subset I.$$

Then the proof of ((2) $\Rightarrow$ (1)) follows from the proof of ((2) $\Rightarrow$ (1)) in Theorem 2.5.8. □

Corollary 2.5.10. *The formula* (2.194) *for the third property (it is called the heat-bath property) in Theorem 2.5.8 provides the precise formulation of what a heat bath is in probabilistic terms when the heat bath Y_n converges to Y on the scale corresponding to Theorem 2.3.8; Similarly, the formula* (2.205) *for the third property in Theorem 2.5.9 provides the precise formulation of what a heat bath is in probabilistic terms when the heat bath Y_n converges to a constant μ on the scale corresponding to Theorem 2.3.9. Through these formulations and the equivalence of the three properties: (1) the subinterval invariant property (2) the invariant temperature property (3) the heat-bath property, we really define an invariant temperature bath mathematically.*

2.6 Appendix

2.6.1 Properties of the large deviation rate function

We include the following properties from [38]. Let $\mathcal{L}$ be the law of X_1, let $\mu := \mathbb{E}[X_1]$ and $\sigma^2 := \text{Var}(X_1)$ and assume that $\sigma > 0$. Let

$$y_- := \inf(\text{supp}(\mathcal{L})), \quad y_+ := \sup(\text{supp}(\mathcal{L}))$$

and ϕ be the function defined in Theorem 2.2.3. Define

$$\mathcal{D}_\phi := \{y \in \mathbb{R} : \phi(y) < \infty\} \quad \text{and} \quad \mathcal{U}_\phi := int(\mathcal{D}_\phi). \tag{2.209}$$

Then the following holds:

1. $\phi(y)$ is convex and lower semi-continious.

2. $0 \leq \phi(y) \leq \infty$ for all $y \in \mathbb{R}$.

3. $\phi(y) = 0$ if and only if $y = \mu$.

4. $\mathcal{U}_\phi = (y_-, y_+)$ and $\phi(y)$ is infinitely differentiable on $\mathcal{U}_\phi$.

5. $\phi''(y) > 0$ on $\mathcal{U}_\phi$ and $\phi''(\mu) = 1/\sigma^2$.

2.6.2 Proof of Corollary 2.2.6

Proof. (2.25) follows from Theorem 2.2.5 since $Z_n \to G$ in distribution and $W_n \to 0$ in probability. (2.26) basically follows from the proof for Berry-Esseen Theorem (see for example Theorem 2.2.8. in [190]). We include a sketch of the proof here.

Let ϕ_Y be the charateristic function of a random variable Y and $\epsilon = \mathbb{E}|X|^3/\sqrt{n}$. To prove (2.26), following every step in the proof given in [190], it sufficies to show that

$$\int_{|t|<c/\epsilon} \frac{|\phi_{\tilde{Z_n}}(t) - \phi_G(t)|}{1+|t|} dt = O(\epsilon), \tag{2.210}$$

for some small constant c. We can show that

$$\begin{aligned}
\left|\phi_{\tilde{Z_n}}(t) - \phi_G(t)\right| &= \left|\exp\left[\frac{-t^2}{2}\left(\frac{n+k}{n}\right) + O\left(\epsilon|t|^3\left(\frac{n+k}{n}\right)\right)\right] - \exp(-t^2/2)\right| \\
&= O\left(\frac{t^2}{n}\exp(-t^2/4)\right) + O\left(\epsilon|t|^3\exp(-t^2/4)\right).
\end{aligned} \tag{2.211}$$

Inserting this to (2.210), after integration, the first term in (2.211) has order $O(\frac{1}{n})$ and the second term has order $O(\epsilon)$. It completes the proof. $\square$

2.6.3 *Proof of the boundedness of Equation* (2.170)

Denote that

$$A(y) := \frac{\partial^2 \log P\big(Y \in [y, y+\delta]\big)}{\partial y^2} = \frac{f_Y'(y+\delta) - f_Y'(y)}{P\big(Y \in [y, y+\delta]\big)} - \left(\frac{f_Y(y+\delta) - f_Y(y)}{P\big(Y \in [y, y+\delta]\big)}\right)^2. \tag{2.212}$$

We can recognize that

$$A(h - \hat{\alpha}_n x) = 2q_n(x),$$

in which the function $q_n(x)$ is defined in Equation (2.92) for the proof of Theorem 2.3.1.

In the entire proof of Theorem 2.3.1, the only place that we use the condition (2.36) regarding uniformly bounded $A(y)$ when $y \in \mathbb{R}$ is just for the proof of Equation (2.93) to show that $\exp(-\psi(I)x) \cdot q_n(x)$ is uniformly bounded on $x \in \mathbb{R}^+$. Therefore, instead of proving uniformly bounded $A(y)$ in the condition (2.36), it suffices to show the uniform boundedness of $\exp(-\psi(I)) \cdot q_n(x)$: there exists a constant C such that

$$|\exp(-\psi(I)x) \cdot A(h - \hat{\alpha}_n x)| \leq C, \quad \hat{\alpha}_n \in (0,1), \quad \text{for all } x \in \mathbb{R}^+. \tag{2.213}$$

By the mean value theorem and the formula of the density of normal distribution, we can show that there exists $\hat{y}$, $\hat{y} \in (y, y+\delta)$ such that the first term on the right side of (2.212) can be written as

$$\begin{aligned}\frac{f_Y'(y+\delta) - f_Y'(y)}{P\big(Y \in [y, y+\delta]\big)} &= (y+\delta)\exp\left[\frac{-y^2+\hat{y}^2}{2\sigma^2}\right]\left(\exp\left[\frac{-2y\delta-\delta^2}{2\sigma^2}\right] - 1\right) + \delta\exp\left[\frac{-y^2+\hat{y}^2}{2\sigma^2}\right] \\ &= (y+\delta)\exp\left[\frac{(\hat{y}-y-\delta)(\hat{y}+y+\delta)}{2\sigma^2}\right] - y\exp\left[\frac{(\hat{y}-y)(\hat{y}+y)}{2\sigma^2}\right].\end{aligned} \tag{2.214}$$

Recall $y < \hat{y} < y+\delta$. When $\hat{y}+y+\delta \in [0, 2h+\delta]$, (2.214) is uniformly bounded. When $\hat{y}+y+\delta < 0$, we can further have

$$\left|\frac{f_Y'(y+\delta) - f_Y'(y)}{P\big(Y \in [y, y+\delta]\big)}\right| \leq (h+\delta)\exp\left[\frac{-\delta(2y+\delta)}{2\sigma^2}\right] + h\exp\left[\frac{-\delta y}{\sigma^2}\right].$$

Therefore

$$\exp(-\psi(I)x)\left|\frac{f_Y'(y+\delta) - f_Y'(y)}{P\big(Y \in [y, y+\delta]\big)}\right| \leq [(h+\delta)\exp(-\delta^2/2\sigma^2) + h] \cdot \exp\left[\frac{-\delta y}{\sigma^2} - \psi(I)x\right]. \tag{2.215}$$

By plugging in $y = h - \hat{\alpha}_n x$, $\hat{\alpha}_n \in (0,1)$ in (2.215), since we have $2\delta/\sigma^2 < \psi(I)$ from the assumptions in Theorem 2.5.3 , we can check the terms on the right hand side in (2.215) is uniformly bounded when $x \in \mathbb{R}^+$.

The second term on the right side of (2.212) can be written as

$$\left(\frac{f_Y(y+\delta) - f_Y(y)}{P\big(Y \in [y, y+\delta]\big)}\right)^2 = \exp\left[\frac{-y^2 + \hat{y}^2}{\sigma^2}\right]\left(\exp\left[\frac{-2y\delta - \delta^2}{2\sigma^2}\right] - 1\right)^2. \tag{2.216}$$

When $y + \hat{y} \in [0, 2h + \delta]$, the right hand side above is uniformly bounded. When $y + \hat{y} < 0$, from (2.216) we have

$$\begin{aligned}
\exp(-\psi(I)x)\left(\frac{f_Y(y+\delta) - f_Y(y)}{P\big(Y \in [y, y+\delta]\big)}\right)^2 &\le \exp\left[\frac{(\hat{y} - y)(\hat{y} + y)}{\sigma^2} - \psi(I)x\right]\left(\exp\left[\frac{-2y\delta - \delta^2}{2\sigma^2}\right] - 1\right)^2 \\
&\le \exp(-\psi(I)x)\left(\exp\left[\frac{-2y\delta - \delta^2}{2\sigma^2}\right] - 1\right)^2 \\
&= \exp(-\psi(I)x)\left(\exp\left[\frac{-2y\delta - \delta^2}{\sigma^2}\right] - 2\exp\left[\frac{-2y\delta - \delta^2}{2\sigma^2}\right] + 1\right).
\end{aligned} \tag{2.217}$$

By plugging in $y = h - \hat{\alpha}_n x$, $\hat{\alpha}_n \in (0,1)$ in (2.217), since we have $2\delta/\sigma^2 < \psi(I)$, we can check the terms on the right hand is uniformly bounded when $x \in \mathbb{R}^+$. Therefore, combining the estimates in two parts, (2.213) is uniformly bounded for all $x \in \mathbb{R}^+$.

2.6.4 Proof of Corollary 2.3.4, Corollary 2.3.10, and Corollary 2.3.12

Proof of Corollary 2.3.4

This proof basically follows the proof in Section 2.4.1 for Theorem 2.3.1, so we only provide the details of the difference here. For the derivation of Equation 2.74, we do Taylor's expansion with respect to x and y for this corollary, so we will get Equations (2.75) - (2.77) as following:

$$\phi_n(I) = \frac{\partial \log P\big(Y_n \in [y, y+\delta] \mid X_n = 0\big)}{\partial y}\bigg|_{y=h} - \frac{\partial \log P\big(Y_n \in [y, y+\delta] \mid X_n = 0\big)}{\partial x}\bigg|_{y=h},$$

$$\begin{aligned} r_n(x) =& \frac{1}{2}\frac{\partial^2 P\big(Y_n \in [y, y+\delta] \mid X_n = \xi\big)}{\partial y^2}\bigg|_{y=h-\alpha_n x,\ \xi=\alpha_n x} \\ &+ \frac{1}{2}\frac{\partial^2 P\big(Y_n \in [y, y+\delta] \mid X_n = \xi\big)}{\partial \xi^2}\bigg|_{y=h-\alpha_n x,\ \xi=\alpha_n x} \\ &- \frac{\partial^2 P\big(Y_n \in [y, y+\delta] \mid X_n = \xi\big)}{\partial \xi \partial y}\bigg|_{y=h-\alpha_n x,\ \xi=\alpha_n x}. \end{aligned} \tag{2.218}$$

With the remaining term

$$k_n(x) = \frac{r_n(x)}{P\big(Y_n \in [h, h+\delta] \mid X_n = 0\big)} - \frac{\phi_n(I)^2 e^{-\gamma_n \cdot \phi_n(I) x}}{2},$$

for some $\alpha_n, \gamma_n \in (0,1)$. Then we obtain

$$f_{X_n|Z_n}(x; I) = \frac{f_{X_n}(x) P(Y_n \in I \mid X_n = 0)(e^{-\phi(I)x} + k_n(x)x^2)}{P(Z_n \in I)}, \quad \text{for } x \in \mathbb{R}^+. \tag{2.219}$$

Based on these new expressions of Equations (2.75) - (2.77), Equation (2.83) becomes

$$\begin{aligned} &\left| \int_{\mathbb{R}^+} A_n f_{X_n}(x) e^{-\phi_n(I)} \log\left(\frac{f_{X_n}(x) P_{Y_n|X_n}\big(I - x; x\big)}{P_{Z_n}\big(I\big)} \cdot \frac{1}{A_n f_{X_n}(x) e^{-\phi_n(I)x}} \right) dx \right| \\ =& \left| \int_{\mathbb{R}^+} A_n f_{X_n}(x) e^{-\phi_n(I)} \log\left(\frac{P_{Y_n|X_n}\big(I - x; x\big)}{P_{Y_n|X_n}\big(I; 0\big) e^{-\phi_n(I)x}} \cdot \frac{P_{Y_n|X_n}\big(I; 0\big)}{P_{Z_n}\big(I\big) A_n} \right) dx \right| \\ \leq& \left| \log\left(\frac{B_n}{A_n}\right) \right| + \left| \int_{\mathbb{R}^+} A_n f_{X_n}(x) e^{-\phi_n(I)} \log\left(\frac{P_{Y_n|X_n}\big(I - x; x\big)}{P_{Y_n|X_n}\big(I; 0\big) e^{-\phi_n(I)x}} \right) dx \right|, \end{aligned} \tag{2.220}$$

where

$$A_n := \frac{1}{\int_{\mathbb{R}^+} f_{X_n}(x) e^{-\phi(I)x} dx} \quad \text{and} \quad B_n := \frac{P_{Y_n|X_n}\big(I; 0\big)}{P_{Z_n}\left(I\right)}. \tag{2.221}$$

From the expression of $f_{X_n|Z_n}\big(x; I\big)$ in (2.219), we have the following identity

$$1 = \int_{\mathbb{R}^+} f_{X_n|Z_n}\big(x; I\big) dx = \frac{B_n}{A_n} + B_n \int_{\mathbb{R}^+} f_{X_n}(x) k_n(x) x^2 dx. \tag{2.222}$$

Equation (2.222) implies

$$\log\left(\frac{B_n}{A_n}\right) = \log\left(1 - B_n\int_{\mathbb{R}^+} f_{X_n}(x)k_n(x)x^2dx\right). \tag{2.223}$$

Now it remains to show

$$\left|B_n\int_{\mathbb{R}^+} f_{X_n}(x)k_n(x)x^2dx\right| \tag{2.224}$$

is small for large n.

By the conditions in Corollary 2.3.4, $P_{Z_n}(I) \geq \delta_2 > 0$, hence there exists a constant $M_1 > 0$ such that

$$B_n = \frac{P_{Y_n|X_n}(I;0)}{P_{Z_n}(I)} \leq M_1. \tag{2.225}$$

And since $k_n(x)$ is uniformly bounded as proof 2.4.1 for Theorem 2.3.1, with the assumption $\mathbb{E}[X_n^2] = a_n$, we can derive that

$$\left|B_n\int_{\mathbb{R}^+} f_{X_n}(x)k_n(x)x^2dx\right| \leq M_1 \cdot \sup|k_n(x)| \cdot \mathbb{E}[X_n^2] = O(a_n). \tag{2.226}$$

Recall (2.223), since $\log(1+x) \leq x$ for all $x > -1$, for sufficiently large n, we have

$$\log\left(\frac{B_n}{A_n}\right) = \log\left(1 - B_n\int_{\mathbb{R}^+} f_{X_n}(x)k_n(x)x^2dx\right) \leq B_n\int_{\mathbb{R}^+} f_{X_n}(x)k_n(x)x^2dx = O(a_n), \tag{2.227}$$

which gives us that the first term in (2.219) is in order $O(a_n)$.

The second term in (2.220) is also in order $O(a_n)$ which follows from the steps (2.92) - (2.94) in Section 2.4.1. Therefore, by the definition of KL-divergence (2.20) and Bayes' theorem for conditional probability and the inequality (2.220), we finally obtain

$$\begin{aligned}
D_{\mathrm{KL}}\left(\hat{\mathbb{P}}_I^{(n)} \parallel \mathbb{Q}_I^{(n)}\right) &= \left|\int_{\mathbb{R}^+} A_n f_{X_n}(x)e^{-\phi_n(I)}\log\left(\frac{f_{X_n|Z_n}(x;I)}{A_n f_{X_n}(x)e^{-\phi_n(I)x}}\right)dx\right| \\
&= \left|\int_{\mathbb{R}^+} A_n f_{X_n}(x)e^{-\phi_n(I)}\log\left(\frac{f_{X_n}(x)P_{Y_n|X_n}(I-x;x)}{P_{Z_n}(I)} \cdot \frac{1}{A_n f_{X_n}(x)e^{-\phi_n(I)x}}\right)dx\right| \\
&= O(a_n).
\end{aligned} \tag{2.228}$$

Proof of Corollary 2.3.10 and Corollary 2.3.12

For the proof of Corollary 2.3.10, since $\log G(I;0) = 0$ and $\log G(I;\xi) \in C(\mathbb{R}^+)$ with respect to ξ, we can do Taylor's expansion for it at zero to get

$$\log G(I;\beta_n x) = \left.\frac{\partial \log G(I;\xi)}{\partial \xi}\right|_0 \beta_n x + O(\beta_n^2 x^2) \quad \text{for } x \in \mathbb{R}^+. \tag{2.229}$$

And similarly, for the proof of Corollary 2.3.12, since $\log R(I;0) = 0$ and $\log R(I;\xi) \in C(\mathbb{R}^+)$ with respect to ξ, we can do Taylor's expansion for it at zero to get

$$\log \left(R(I;\beta_n x)\right)^{\frac{1}{\beta_n}} = \frac{1}{\beta_n}\left(\left.\frac{\partial \log R(I;\xi)}{\partial \xi}\right|_0 \beta_n x + O(\beta_n^2 x^2)\right) \quad \text{for } x \in \mathbb{R}^+. \tag{2.230}$$

Then the proof of Corollary 2.3.10 follows from the proof given in Section 2.4.2 with an additional linear term $\frac{\partial \log G(I;0)}{\partial \xi}\beta_n x$ in (2.140); and the proof of Corollary 2.3.12 follows from the proof in Section 2.4.2 with an additional linear term $\frac{\partial \log R(I;0)}{\partial \xi} x$ in (2.141).

Chapter 3

GENERALIZING GIBBSIAN STATISTICAL ENSEMBLE THEORY FOR STRONGLY COUPLED HETEROGENEOUS SYSTEMS

This chapter is based on joint work with Hong Qian and Wenning Wang [32].

3.1 Introduction

Gibbs' theory proper concerns with finite size systems that are in contact with a *temperature bath*. It reproduces the classical thermodynamic results as the limiting behavior when the size tending to infinity, e.g., *thermodynamic limit*. The Laplace transform of partition functions becomes the Legendre transform of thermodynamic potentials in the limit. The theory has been widely checked against experiments on finite and even small systems of liquids, solutions, and macromolecules. In biophysical chemistry, it has been used as a natural law for further understanding molecular interactions [178].

Through a rigorous mathematical analysis, we recently argued [31] that not only the macroscopic thermodynamics is an emergent phenomenon, the very Gibbs' formalism itself is a result of a limit law according to the theory of *conditional probability*. The limit law is to the canonical distribution what the central limit theorem is to the Gaussian fluctuation theory developed by Einstein and Landau [119]. And as in the theories of phase transition [3] and the passing from quantum mechanics to quantum chemistry via Born-Oppenheimer approximation, the mathematical limit involves subtleties that have fundamental importance [33].

The aim of this paper is to generalize the Gibbs' theory to strongly coupled, heterogeneous systems: Consider a system $\mathfrak{S}$ coupled to its bath $\mathfrak{W}$ with a setup: (i) interactions between $\mathfrak{S}$ and $\mathfrak{W}$ might not be negligible; (ii) $\mathfrak{S}$ can be distinct to $\mathfrak{W}$ with respect to the physical properties, e.g., water with ice floating, or the chemical composition, e.g., side-chain conformational variations in protein structure [212] and a RNA molecule immersed in aqueous solution [97]. However, current mathematical theories, *the equivalence of ensembles* [108, 133, 39, 123, 77] and *the Gibbs conditioning principle* [217, 196, 188, 38], are not sufficient to justify the state space density of $\mathfrak{S}$ for the following reasons: The former fails in certain circumstance when the system has critical phenomena [133, 195, 66], and the later requires $\mathfrak{W}$ as a large number of identical copies of $\mathfrak{S}$.

To logically solve this issue, a new result is presented in this work: Let U_1 and U_2 be the bare energy of $\mathfrak{S}$ and $\mathfrak{W}$, respectively, and U_{12} be the interaction energy. Now, the distribution of U_1, given a conserved total energy $U_t = U_1 + U_{12} + U_2$, can be expressed by the conditional probability $f_{U_1|U_t}$. As $U_1/U_t \to 0$, $f_{U_1|U_t}$is asymptotic to f_{U_1} weighted by an exponential factor whose exponent is determined by both of the fluctuations of U_2 and the correlation caused by U_{12}. This new result is based on the key lemma provided in the next section. Under the assumption of the *principle of equal a priori probabilities* [92], we further apply that lemma to derive the state space density of $\mathfrak{S}$.

In addition, having that key lemma also provides a lens for helping us look through the grand canonical ensemble from a different perspective: Consider a group of particles distributed randomly in a space. Let K and L be the particle numbers in two adjacent regions of the space, and the region of K be infinitesimal in comparison to the region of L. Then the lemma implies that conditional distribution $P_{K|K+L}$ is asymptotic to P_K weighted by an exponential factor whose exponent is solely determined by the fluctuations of L. From this perspective, undetermined P_K yields the freedom of choosing a prior distribution, which echoes the "subjectivity" in the informatic view of resolving the Gibbs paradox [27].

3.2 Three mathematical theorems

In the following lemma, X is indexed by n to represent an infinitesimal system and no assumption of independence between "system of interest" X and "heat bath" Y is made. Thm 1. shows however the importance of X, Y independence in defining a unambiguous equilibrium temperature for all Xs that are in contact with a given Y. Thm 2. is an extension of the lemma to integer random variables. Thm 3. further clarifies the condition under which the temperature is uniquely defined for a heat bath irrespective of the details of the "contact". We state the key results without the proofs, see [31] for the full mathematical details.

New result of applying the Lemma to strongly coupled systems is presented in a later section.

Lemma, a limit law for a sequence of conditional probabilities: Consider a sequence of non-negative random variable $X_n \in \mathbb{R}$ and another non-negative random variable $Y \in \mathbb{R}$ which needs not to be independent of X_n, on a probability space $(\Omega, \mathcal{F}, P)$. Denoting $H_n = X_n + Y$ and $\mathbb{E}[X_n^2] = a_n^2$, where $a_n \to 0$ as $n \to \infty$. Then the sequence of conditional probability density functions has an asymptotic expression [1]:

$$f_{X_n|H_n}(x; h_\delta) \simeq Z_n^{-1}(h_\delta) f_{X_n}(x) e^{-\psi_n(h_\delta)x}, \tag{3.1a}$$

as $n \to \infty$, where $h_\delta := [h, h+\delta] \in \mathbb{R}$, $f_{X_n}(x)$ is the marginal distribution of the correlated pair (X_n, Y), Z_n is the normalization factor, and

$$\psi_n(h_\delta) = \left.\frac{\partial \ln P(Y \in y_\delta)}{\partial y}\right|_{y=h} \tag{3.1b}$$

$$+ \left[\frac{\partial \ln C_n(y_\delta; x)}{\partial y} - \frac{\partial \ln C_n(y_\delta; x)}{\partial x}\right]_{x=0, y=h}, \tag{3.1c}$$

where $y_\delta := [y, y+\delta]$ and

$$C_n(y_\delta; x) = \frac{P(Y \in y_\delta | X_n = x)}{P(Y \in y_\delta)}. \tag{3.1d}$$

If Y is independent from X_n, $C_n = 1$ and $\psi_n(h_\delta)$ is independent of n; it is solely determined by the property of $P(Y \in [y, y+\delta])$ near $y = h$. In the following theorems, $(\partial \ln P(Y \in y_\delta)/\partial y)_{y=h}$ is denoted by $\beta^*(h_\delta)$.

Theorem 1, canonical Gibbs distribution on phase space. Consider a compact continuous phase space $\Omega \subset \mathbb{R}^{2n_1+2n_2}$ with state variable $(\mathbf{w}_1, \mathbf{w}_2)$, $\mathbf{w}_1 \in \mathbb{R}^{2n_1}$, $\mathbf{w}_2 \in \mathbb{R}^{2n_2}$. Furthermore $\mathbf{w}_1$ and $\mathbf{w}_2$ are independent w.r.t. P. Assume three non-negative continuous functions that satisfy $U_t(\mathbf{w}_1, \mathbf{w}_2) = U_1(\mathbf{w}_1) + U_2(\mathbf{w}_2)$ and $U_1(\mathbf{w}_1) \ll U_2(\mathbf{w}_2)$. Let $Y := U_2(\mathbf{w}_2)$, then the limit distribution in (3.1a) implies

[1] The two probability density functions on lhs and rhs, in term of their KL divergence, has the order of $O(a_n)$.

$$P(U_1(\mathbf{w}_1) \leq a | U_t(\mathbf{w}_1, \mathbf{w}_2) \in h_\delta)$$
$$\simeq \int_{U_1(\mathbf{w}_1)\in[0,a]} Z^{-1}(\beta^*) e^{-\beta^* U_1(\mathbf{w}_1)} \mu_1(\mathrm{d}\mathbf{w}_1), \tag{3.2a}$$

in which μ_1 is the induced measure of P by the projection from $(\mathbf{w}_1, \mathbf{w}_2)$ to $\mathbf{w}_1$. Under the assumption of the principle of equal a priori probabilities for all $\mathbf{w}_1$ in each set $U_1(\mathbf{w}_1) \in [a, a + \mathrm{d}a]$, then the conditional density of $\mathbf{w}_1$ follows

$$f(\mathbf{w}_1 | U_t \in h_\delta) \simeq Z^{-1}(\beta^*) e^{-\beta^* U_1(\mathbf{w}_1)}. \tag{3.2b}$$

Theorem 2, counting statistics and grand canonical distribution. Consider a sequence of non-negative integer random variables $L_n \in \mathbb{Z}$, and another non-negative integer random variable $K \in \mathbb{Z}$ which is independent to L_n. Denoting $M_n = K + L_n$. There exists $b_n \to \infty$ and $c_n \geq 0$ such that $\tilde{L}_n := (L_n - c_n)/b_n$ converges to a continuous random variable $\tilde{Y}$ in distribution. Then the sequence of conditional probability mass functions has an asymptotic expression [2]:

$$p_{K|M_n}(k; b_n h_\delta + c_n) \simeq Q_n^{-1}(h_\delta) p_K(k) e^{\mu(h_\delta)k/b_n}, \tag{3.3a}$$

as $n \to \infty$, $p_K(k)$ is the marginal distribution of K, Q_n is the normalization factor, and

$$\mu(h_\delta) = -\frac{\partial \ln P(\tilde{Y} \in y_\delta)}{\partial y}\bigg|_{y=h}. \tag{3.3b}$$

Furthermore, the product of (3.2b) and (3.3a) yields the joint probability of $\mathbf{w}_1$ and K conditioned on both U_t and M_n for the grand canonical distribution.

The interval $h_\delta := [h, h+\delta]$ plays an essential role in the forementioned theorems; it determines the exponent for the exponential (or geometric) factor. Thm. 3 illustrates that the above results become minimally dependent upon the choice of h_δ if the $P(Y \in h_\delta)$ satisfies certain invariant criteria.

Theorem 3, invariance w.r.t. choosing h_δ: Under the same setup as Lemma with Y independent of X_n, let us consider subinterval $h'_{\delta'} := [h', h' + \delta'] \subset h_\delta$. The following three statements are equivalent:

[2]The two probability functions on lhs and rhs, in term of their KL divergence, has the order of $O(b_n^{-1})$.

1. Invariance w.r.t. subintervals: For all subintervals $h'_{\delta'}$, the distance in terms of the total variation between $f_{X_n|H_n}(x;h'_{\delta'})$ and $f_{X_n|H_n}(x;h_\delta)$ converges to zero as $n\to\infty$.

2. Equal exponent: For all subintervals $h'_{\delta'}$,

$$\left.\frac{\partial \ln P(Y\in y_{\delta'})}{\partial y}\right|_{y=h'} \equiv \beta^*(h_\delta). \tag{3.4a}$$

3. The heat-bath property: The conditional probability density for Y at $y\in[h,h+\delta]$ is itself exponentially distributed as:

$$\begin{aligned} f_{Y|\{Y\in h_\delta\}}(y) &= \lim_{\delta'\to 0}\frac{1}{\delta'}P(Y\in y_{\delta'}|Y\in h_\delta) \\ &= \frac{\beta^* e^{\beta^* y}}{e^{\beta^*(h+\delta)}-e^{\beta^* h}}. \end{aligned} \tag{3.4b}$$

3.3 New insights from the theorems

Dynamic conservation vs. conditional probability — Let us consider a *Gedankenexperiment* on a very large mechanical system $\mathfrak{S}\cup\mathfrak{W}$ in which the small part $\mathfrak{S}$ has mechanical energy fluctuations. If one only measures the *static statistics* of the energy of $\mathfrak{S}$, should the resulting statistics be different between (a) $\mathfrak{S}\cup\mathfrak{W}$ has a conserved total mechanical energy $E_{\mathfrak{S}\cup\mathfrak{W}}\in[h,h+\delta]$, with infinitesimal δ, for all time determined by the initial condition for the dynamics; and (b) $\mathfrak{S}\cup\mathfrak{W}$ has a fluctuating total mechanical energy $E_{\mathfrak{S}\cup\mathfrak{W}}(t)$ as a function of time, but one selects only those measurements on $\mathfrak{S}$ that simultaneously has $E_{\mathfrak{S}\cup\mathfrak{W}}\in[h,h+\delta]$?

If one installs an equivalent principle between (a) and (b), the lemma can be interpreted as follows: First, the fixed Y with varying X indexed by n implies that the system of interest is relatively small to its heat bath, and it is called the *heat-bath limit* as $n\to\infty$. Second, the total energy $X+Y=h$ is conserved, subjected to small fluctuations δ. This is understood as the *microcanonical ensemble*. Third, in particular, if the system is asymptotically uncorrelated with its heat bath in the heat-bath limit: This is reflected in the correlation-related term (3.1d) converging to 0 as $n\to\infty$. Then it arises as a universal parameter $\beta^*(h_\delta)=\partial\ln P(Y\in h_\delta)/\partial y$ which is solely determined by the fluctuations of the heat bath.

Independent heterogeneous systems **—** In thm. 1, Ω represents the compact, continuous phase space of the total system with microstate variable $(\mathbf{w}_1, \mathbf{w}_2)$. Let Ω_1 be the phase space for $\mathbf{w}_1$ and Ω_2 be the phase space for $\mathbf{w}_2$. Then the structure of Thm. 1 is as follows:

$$(\Omega, \mathcal{B}(\Omega), P) \xrightarrow{\pi_1} (\Omega_1, \mathcal{B}(\Omega_1), \mu_1) \xrightarrow{U_1} (\mathbb{R}, \mathcal{B}(\mathbb{R}), F_{U_1}),$$
$$(\Omega, \mathcal{B}(\Omega), P) \xrightarrow{\pi_2} (\Omega_2, \mathcal{B}(\Omega_2), \mu_2) \xrightarrow{U_2} (\mathbb{R}, \mathcal{B}(\mathbb{R}), F_{U_2}),$$

in which π_1, π_2 are the projection maps with the corresponding induced measures μ_1, μ_2, and U_1, U_2 are the functions for observables with the distributions F_{U_1}, F_{U_2}. To begin with the lemma, we have $F_{U_1|U_1+U_2}(x; h_\delta)$ at the level of observables; furthermore, by the principle of equal a priori probabilities for all $\mathbf{w}_1$ in each thin shell of U_1, we then obtain $f(\mathbf{w}_1|U_1 + U_2 \in h_\delta)$ at the level of microstates. The principle of equal a priori probabilities can be justified by

$$\begin{aligned} & f(\mathbf{w}_1|U_1(\mathbf{w}_1) \in \mathrm{d}a, U_t \in h_\delta) \\ =& f(\mathbf{w}_1|U_1(\mathbf{w}_1) \in \mathrm{d}a) \propto \frac{\mathbf{1}_{\{U_1(\mathbf{w}_1)\in \mathrm{d}a\}}}{P(U_1(\mathbf{w}_1) \in \mathrm{d}a)}, \end{aligned} \tag{3.6}$$

where the the first equation is due to independent (U_1, U_2) and then $f(\mathbf{w}_1|U_1(\mathbf{w}_1) \in \mathrm{d}a)$ has to be uniform based on the ergodic theory that all microstates $\mathbf{w}_1$ confined to each thin energy shell $U_1(\mathbf{w}_1) \in \mathrm{d}a$ should be equally probable.

The strength of this approach can be seen when it is applied to heterogeneous systems: Let us consider the example for a molecule immersed in an aqueous solution with negligible interactions. Now, $\mathbf{w}_1$ represents the microstate of the molecule and $\mathbf{w}_2$ represents the microstate of the solution. In this case, since the molecule is distinct to the aqueous solution, so a justification of the distribution of $\mathbf{w}_1$ was missing in the previous mathematical theories based on a homogeneous total system. On the other hand, by applying the lemma, one only need to check $U_1(\mathbf{w}_1) \ll U_2(\mathbf{w}_2)$ to obtain the distribution of $U_1(\mathbf{w}_1)$ in Eq. (3.2a); Furthermore, given the principle of equal a priori probabilities for $\mathbf{w}_1$[3], the density of $\mathbf{w}_1$ in Eq. (3.2b) can be further justified.

---

[3]If all the potential and kinetic energy of a molecule are considered, then this is a mechanical system, which follows the ergodic theory.

Strongly coupled heterogeneous systems — Following from the above example, here we consider the molecule has strong interactions with the solution, hence the total energy

$$U_t(\mathbf{w}_1, \mathbf{w}_2) = U_1(\mathbf{w}_1) + U_{12}(\mathbf{w}_1, \mathbf{w}_2) + U_2(\mathbf{w}_2), \tag{3.7}$$

in which $U_{12}(\mathbf{w}_1, \mathbf{w}_2)$ is a non-negligible interaction energy. If we simply group $U_{12}(\mathbf{w}_1, \mathbf{w}_2)$ and $U_2(\mathbf{w}_2)$ together into a new observable $Y := U_{12}(\mathbf{w}_1, \mathbf{w}_2) + U_2(\mathbf{w}_2)$. Now Y and $X := U_1(\mathbf{w}_1)$ are no longer independent. Parallel to Thm. 1, via the lemma, we still have $F_{X|X+Y}(x; h_\delta)$ at the level of observables. However, since (X, Y) is non-independent, the first equation in (3.6) is no longer true. Thus the principle of equal a priori probabilities for $\mathbf{w}_1$ might be invalid and we are not be able to further derive the density of $\mathbf{w}_1$.

Let us consider a "new Hamiltonian" $\hat{U}_1$ such that

$$f(\mathbf{w}_1 | \hat{U}_1(\mathbf{w}_1) \in \mathrm{d}a, U_t \in h_\delta) \propto \frac{\mathbf{1}_{\{\hat{U}_1(\mathbf{w}_1) \in \mathrm{d}a\}}}{P(\hat{U}_1(\mathbf{w}_1) \in \mathrm{d}a)}, \tag{3.8}$$

and assume the existence of this Hamiltonian for a while (we will discuss it later). Now, despite the redefined pair of observables $\hat{X} := \hat{U}_1(\mathbf{w}_1)$ and $\hat{Y} := U_t(\mathbf{w}_1, \mathbf{w}_2) - \hat{U}_1(\mathbf{w}_1)$ still dependent, since Eq. (3.8) guarantees $\mathbf{w}_1$ is uniform in each shell $\hat{U}_1(\mathbf{w}_1) \in \mathrm{d}a$, we can further obtain the density of $\mathbf{w}_1$

$$f^{(m)}(\mathbf{w}_1 | U_t \in h_\delta) \propto e^{-\psi(h_\delta)\hat{U}_1(\mathbf{w}_1)}. \tag{3.9}$$

The superscript m indicates the conditional density as a result of the assumption that the isolated total system obeys the microcanonical ensembles.

Recently, Jarzynski [97], Talkner and Hänggi [189] have discussed strongly coupled systems within a total system that is canonically distributed:

$$f^{(c)}(\mathbf{w}_1, \mathbf{w}_2) \propto \exp\{-\gamma\left[U_1(\mathbf{w}_1) + U_{12}(\mathbf{w}_1, \mathbf{w}_2) + U_2(\mathbf{w}_2)\right]\}. \tag{3.10}$$

The marginal density then

$$f^{(c)}(\mathbf{w}_1) \propto e^{-\gamma\left[U_1(\mathbf{w}_1) + \phi^{(c)}(\mathbf{w}_1;\gamma)\right]}, \tag{3.11}$$

$$\phi^{(c)}(\mathbf{w}_1; \gamma) = -\frac{1}{\gamma} \ln \frac{\int \mathrm{d}\mathbf{w}_2 e^{-\gamma[U_{12}(\mathbf{w}_1,\mathbf{w}_2) + U_2(\mathbf{w}_2)]}}{\int \mathrm{d}\mathbf{w}_2 e^{-\gamma U_2(\mathbf{w}_2)}},$$

where $U_1(\mathbf{w}_1)+\phi^{(c)}(\mathbf{w}_1;\gamma)$ is often called the Hamiltonian of mean force [97, 189]; As the function depends only on the positions, it matches Kirkwood's potential of mean force [110]; the force is acting on $\mathbf{w}_1$ and averaged over all the $\mathbf{w}_2 \in \mathbb{R}^{2n_2}$.

As the heat bath is infinitely large in relative to the system of interest, let us set a pair of equations

$$\begin{aligned} &\psi(h_\delta) = \gamma, \\ &\hat{U}_1(\mathbf{w}_1) = U_1(\mathbf{w}_1) + \phi^{(c)}(\mathbf{w}_1;\gamma) \end{aligned} \tag{3.12}$$

and therefore $f^{(m)} = f^{(c)}$. Based on this setup, we have two significant findings: (i) By the theory of equivalence of ensembles [195], if $f^{(m)} = f^{(c)}$ in the limit, then the reciprocal of temperature γ in the canonical ensemble is uniquely determined by a function of the conserved energy h_δ in the microcanonical ensemble. From the pair of equations (3.12), we have a representation of this function by ψ:

$$\begin{aligned} \psi(h_\delta) = &\left.\frac{\partial \ln P(\hat{Y} \in y_\delta)}{\partial y}\right|_{y=h} \\ &+ \left[\frac{\partial \ln C(y_\delta;x)}{\partial y} - \frac{\partial \ln C(y_\delta;x)}{\partial x}\right]_{x=0,y=h}. \end{aligned} \tag{3.13}$$

As $\delta \to 0$, function $C(y,x)$ is known as the *copula density*

$$C(y,x) = \frac{f_{\hat{Y}|\hat{X}}(y;x)}{f_{\hat{Y}}(y)} = \frac{f_{\hat{Y},\hat{X}}(y,x)}{f_{\hat{Y}}(y) f_{\hat{X}}(x)}, \tag{3.14}$$

and it is intimately related to the *mutual information* of $(\hat{Y},\hat{X})$

$$I(\hat{Y};\hat{X}) = \int_y \int_x f_{\hat{Y},\hat{X}}(y,x) \ln \frac{f_{\hat{Y},\hat{X}}(y,x)}{f_{\hat{Y}}(y) f_{\hat{X}}(x)}. \tag{3.15}$$

(ii) The new Hamiltonian $\hat{U}_1(\mathbf{w}_1)$ given by the pair of equations (3.12) guarantees the principle of equal a priori probabilities for $\mathbf{w}_1$ in each energy shell defined by $\hat{U}_1(\mathbf{w}_1)$. From this result, we can justify that the most reasonable Hamiltonian of the system of interest in a strongly coupled system should be redefined by its Hamiltonian of mean force. But we shall emphasize that the equivalence of ensembles is not always satisfied when a system has critical phenomena. In this situation, there may not exist a function $\hat{U}_1(\mathbf{w}_1)$ satisfying the pair of equations (3.12). Alternatively, while the probabilistic equation (3.8) is relatively vague to the pair of equations (3.12), as long as we can find the solution $\hat{U}_1(\mathbf{w}_1)$ directly from it, then the solution itself has to satisfy the principle of equal a priori probabilities. In other words, Eq. (3.8) should be the most fundamental equation of the Hamiltonian of mean force when the total system is assumed an isolated system (microcanical ensemble). Having the solution of $\hat{U}_1(\mathbf{w}_1)$ from it, state space density given by $f^{(m)}$ is generally true for a system strongly coupled to its bath with total energy conservation, even if the equivalence of ensembles is invalid.

***Grand canonical ensemble and Gibbs paradox* —** Gibbs introduced the notion of a *grand canonical ensemble* for a system that has fluctuating particle numbers due to exchanges with a single material reservoir. Here, Legendre transform inherited from thermodynamics has met with difficulties as a unifying mathematical device for the ensemble theory. The existence of the Gibbs paradox is an indication that a more unified, rigorous treatment is desirable. A recent such attempt based on the principle of maximum relative entropy can be found in [9, 71]. Our probabilistic theory provides a different perspective to understand the Gibbs paradox as follows:

In Thm. 3, if the integer random variables K and L_n are the number of particles *uniformly distributed* in a space within regions $\mathcal{B}$ and $\mathcal{D}\setminus\mathcal{B}$, $\mathcal{B}\subset\mathcal{D}$, and we assume the size of $\mathcal{B}$ is infinitesimal in comparison to the whole space, then prior distribution $P_K(k)$ should follow a Poisson distribution [71]. We see clearly that the $1/k!$ arises through the distribution of K *a priori* for counting particles distributed in space, which is known as spatial Poisson point processes. Changing representation from Lagrangian tracking to Eulerian counting of particles gives rise to the notion of entropy of assimilation [11]: For Lagrangian particles the $k!$ is absent; the freedom of choosing priors corresponds to the “subjective observer” discussed in [27].

***Idealized heat bath* —** Our mathematical theory is based on two elements: (i) a small system that is coupled to a much larger one, and (ii) the total system has an observable, an “energy” function, that is restricted within an interval of infinitesimal fluctuations. A question naturally arises: What are the conditions under which the larger system can be idealized as a heat bath that gives the small system of interest a “unique temperature” that is independent of the details of h and δ? This question is answered by Thm. 3: The heat bath itself has an exponentially distributed energy fluctuation, conditioned on the interval h_δ; the single parameter on the exponent defines the temperature.

It is highly instructive to note that such an exponential behavior can be justified in terms of the large deviations theory for systems with extensive, Eulerian-homogeneous thermodynamic variables [128]. In this case, one identifies the U_t, a thermodynamic potential, with Y having a probability proportional to e^{-ny}, in which $n \to \infty$ indicates the thermodynamic limit. Then the normalized density of Y satisfies Eq. (3.4b). The Y as a bath provides a small system X that is in contact a temperature.

3.4 Discussion

Legendre transform in thermodynamics — With one additional assumption, the result in Eq. (3.1a) a gives rise to the Legendre transform in thermodynamics. If one assumes the existence of a thermodynamic limit, with thermodynamic energies and entropies being "extensive quantities" [4], i.e., $S(x) \equiv \ln f_X(x)$ and $\beta^* x$ are both $\propto\ V$, where $V \to \infty$ representing the size of the system, then the normalization factor in (3.1a)

$$\begin{aligned} F(\beta^*) &\equiv -(\beta^*)^{-1} \ln Z \ = \ -(\beta^*)^{-1} \ln \int_0^{h+\delta} e^{S(x)-\beta^* x} \mathrm{d}x \\ &\simeq \Big[E - (\beta^*)^{-1} S(E) \Big]_{\mathrm{d}S(E)/\mathrm{d}E=\beta^*}. \end{aligned} \tag{3.16a}$$

We note that the thermodynamic equation (3.16a) is for the system of interest, in addition to being small relative to its heat bath, is itself tending a thermodynamic limit with its size $V \to \infty$. In contradistinction to the latter, the $n \to \infty$ in the lemma represents the heat-bath limit: the heat bath is infinitely large in relative to the system. Therefore, if the system and its heat bath are not independent, in both of the thermodynamic and heat-bath limits, the Eq. (3.16a) becomes

$$F(\psi^*) \simeq \Big[E - (\psi^*)^{-1} S(E) \Big]_{\mathrm{d}S(E)/\mathrm{d}E=\psi^*}, \tag{3.16b}$$

$$\psi^* = \beta^* + \gamma^*, \tag{3.16c}$$

[4] This is closely related to the large deviations principle in the theory of probability, and the widely cited "maximum term method" in standard textbook.

where

$$\gamma^* = \lim_{n\to\infty} \left[\frac{\partial \ln C_n(y_\delta; x)}{\partial y} - \frac{\partial \ln C_n(y_\delta; x)}{\partial x} \right]_{x=0, y=h}. \tag{3.16d}$$

In Eqs. (3.16a) and (3.16b), let E_β be the solution of $\mathrm{d}S(E)/\mathrm{d}E = \beta^*$ and E_ψ be the solution of $\mathrm{d}S(E)/\mathrm{d}E = \psi^*$, then $E_\psi - E_\beta$ is the change in the macroscopic energy due to the interaction effect. This change is a consequence of the non-zero γ^* in Eq. (3.16d), which means the system and its heat bath are infinitely strongly correlated in the limits. One can hypothesize that correlations as such are due to long-range interactions. Further mechanistic investigations are required in the future.

Prior knowledge of systems — As shown in Thm. 2, our theory provides the freedom of choosing prior distributions for the grand canonical ensemble. Similarly, the prior distributions are also undetermined in Thm. 1 for the generalized canonical ensemble. In particular, if one chooses the uniform prior distribution for the system of interest, i.e., $\mu_1(\mathrm{d}\mathbf{w}_1) = \mathrm{d}\mathbf{w}_1 / \int_{\Omega_1} \mathrm{d}\mathbf{w}_1$, then Eq. (3.2b) becomes the density of $\mathbf{w}_1$ w.r.t. the *Lebesgue measure*, which is precisely as the canonical ensemble in standard textbooks [120, 92, 147]. Additionally, if one also chooses the uniform prior distribution for the heat bath, i.e., $\mu_2(\mathrm{d}\mathbf{w}_2) = \mathrm{d}\mathbf{w}_2 / \int_{\Omega_2} \mathrm{d}\mathbf{w}_2$, then the parameter of the exponential factor becomes

$$\beta^* = \left. \frac{\partial \ln P(Y \in y_\delta)}{\partial y} \right|_{y=h} = \left. \frac{\partial \ln \Gamma(y)}{\partial y} \right|_{y=h}, \tag{3.17a}$$

where Γ is the volume of a set of phase points

$$\Gamma(y) = \int_{U_2(\mathbf{w}_2) \in [y, y+\delta]} \mathrm{d}\mathbf{w}_2, \tag{3.17b}$$

and $\ln \Gamma$ is corresponding to the Boltzmann's entropy. The reciprocal of (3.17a) is known as the *absolute temperature* in the textbooks. Furthermore, by adding the correlation-related term (3.1c) to Eq. (3.17a), we then show that a strongly coupled heterogeneous system, assumed each part has its own natural structure obeying the uniform priori, has a temperature T^*

$$\frac{1}{T^*} = \left. \frac{\partial \ln \Gamma(y)}{\partial y} \right|_{y=h} + \gamma^*. \tag{3.18}$$

This emergent temperature is contributed by the absolute temperature and a correction term due to infinitely strong correlations in the heat-bath limit.

Chapter 4

COUNTING SINGLE CELLS AND COMPUTING THEIR HETEROGENEITY: FROM PHENOTYPIC FREQUENCIES TO MEAN VALUE OF A QUANTITATIVE BIOMARKER

This chapter is based on joint work with Hong Qian [168].

4.1 Introduction

Statistical analyses of data and stochastic models of mechanisms are two very different, but complementary approaches in biological research. While the former obtains a quantitative representation of high-throughput measurements [150], the latter can provide "laws of nature" through limit theorems [33], widely called *emergent phenomenon*. A case in point is the theory of phase transition [3] which shows that a nonlinear stochastic dynamical system with bistability and cusp catastrophe, in the limit of time $t \to \infty$ followed by system's size $V \to \infty$, necessarily exhibits a discontinuous transition [165]. Another example is the recent work [74] which demonstrates that Gibbsian equilibrium chemical thermodynamics can be reformulated as a limit theorem in a mesoscopic chemical kinetic system, with N species and M reversible stochastic elementary reactions, as the system's size becoming macroscopic.

With the rise of single-cell biology, one naturally is interested in the limiting behavior of the phenotypic frequencies among a population of cells, usually based on one, or several biomarkers. In this case, there is actually a very powerful mathematical result that is widely known to probabilists and statistical physicists. In this tutorial, we give an introduction of this theory and discuss its broader implications.

4.2 Characterizing heterogeneity in single cells

Asymptotic probaility distribution for sample frequencies of cellular phenotypes— To study phenotypic heterogeneity, let a population of N isogenic cells as independent and identically distributed (i.i.d.) realizations of random events from a set $\Omega = \{1, 2, \cdots, M\}$: There are totally M possible phenotypes. Among the N cells, let n_k denotes the random number of cells in the k^{th} state: $n_1 + n_2 + \cdots + n_M = N$. By *phenotypic frequency*, we mean $f_k^{(N)} \equiv \frac{n_k}{N}$.

Let p_k denote the probability of a cell in the k^{th} state. Then the probability distribution for the observed frequency $\vec{f} = (f_1, \cdots, f_M)$ being $\mathbf{x} = (x_1, \cdots, x_M)$ follows a multinomial distribution

$$\Pr\left\{\vec{f}^{(N)} = \mathbf{x}\right\} = \frac{N!}{(Nx_1)!(Nx_2)! \cdots (Nx_M)!} p_1^{Nx_1} p_2^{Nx_2} \cdots p_M^{Nx_M}. \tag{4.1}$$

Since usually N is very large in a high-throughput single-cell experiment, one can safely approximate Eq. (4.1) using Stirling's formula and obtain:

$$\begin{aligned} \ln \Pr\left\{\vec{f}^{(N)} = \mathbf{x}\right\} &= \ln\left(\frac{N!}{(Nx_1)!(Nx_2)! \cdots (Nx_M)!} p_1^{Nx_1} p_2^{Nx_2} \cdots p_M^{Nx_M}\right) \\ &\simeq -N\left[\sum_{k=1}^{M} x_k \ln\left(\frac{x_k}{p_k}\right)\right]. \end{aligned} \tag{4.2}$$

Therefore, one has the asymptotic limit

$$\varphi(\mathbf{x}) = -\lim_{N\to\infty} \frac{1}{N} \ln \Pr\left\{\vec{f}^{(N)} = \mathbf{x}\right\} = \sum_{k=1}^{M} x_k \ln\left(\frac{x_k}{p_k}\right). \tag{4.3}$$

In the theory of large deviations of probability, this is known as Sanov's theorem [38]. Since $\varphi(\mathbf{x}) > 0$ except when $\mathbf{x} = (p_1, \cdots, p_M)$, in the limit of $N \to \infty$, the probability of $\vec{f}_k^{(N)} \neq p_k$ is zero, and the probability of $\vec{f}_k^{(N)} = p_k$ is one. The frequency yields the probability for an infinitely large number of i.i.d. samples. Furthermore, Eq. (4.2) shows that $(p_1, p_2, \cdots, p_M)$ are the most probable sample frequencies for a finite but large N.

Asymptotic distribution for the mean value of a biomarker — Eqs. (4.1) and (4.2) give the probability for the frequencies within the N cells distributed among the M phenotypic states. We now consider a specific biomarker g, which is assumed to be a well defined real-valued function of the phenotype of a cell: $\mathrm{g} = g_k$ when a cell is in the k^{th} state.

It is very clear that if one knows the frequencies $\vec{f}^{(N)}$, then the mean value for g over the entire population of the N cells is determined:

$$\overline{\boldsymbol{\gamma}}^{(N)} = \frac{n_1 g_1 + n_2 g_2 + \cdots + n_M g_M}{N} = \sum_{k=1}^{M} f_k^{(N)} g_k; \tag{4.4}$$

since the frequencies $f_k^{(N)}$ are random, so is $\overline{\mathbf{g}}^{(N)}$. Then when $N \to \infty$, one expectes the $\overline{\mathbf{g}}^{(N)}$ approaching to the expected value $\mathbb{E}[\mathbf{g}]$. This is easy to show:

$$\lim_{N\to\infty} \overline{\mathbf{g}}^{(N)} = \sum_{k=1}^{M} \left(\lim_{N\to\infty} f_k^{(N)} \right) g_k = \sum_{k=1}^{m} p_k g_k = \mathbb{E}[\mathbf{g}]. \tag{4.5}$$

What is the probability distribution for $\overline{\boldsymbol{\gamma}}^{(N)}$ when N is very large but not infinite? One can calculate this:

$$\begin{aligned} \Pr\left\{\overline{\mathbf{g}}^{(N)} = y\right\} &= \sum_{\left\{\mathbf{x}:\, \sum_{k=1}^{M} x_k g_k = y\right\}} \Pr\left\{\vec{f}^{(N)} = \mathbf{x}\right\} \\ &\simeq \exp\left(-N \inf_{\left\{\mathbf{x}:\, \sum_{k=1}^{M} x_k g_k = y\right\}} \varphi(\mathbf{x}) \right) \quad \text{as } N \to \infty. \end{aligned} \tag{4.6}$$

We obtain the (4.6) because among the many sets of $\mathbf{x}$ that give the same value y, each has a probability of $e^{-N\varphi(\mathbf{x})}$. Therefore, as $N \to \infty$, only the set with the smallest $\varphi(\mathbf{x})$ matters. Eq. (4.6) indicates that for very large N, the probability distribution for the mean value of the biomarker $\overline{\mathbf{g}}^{(N)}$ has the form $e^{-N\psi(y)}$, in which

$$\psi(y) \equiv -\lim_{N\to\infty} \frac{1}{N} \ln \Pr\left\{\overline{\mathbf{g}}^{(N)} = y\right\} = \inf_{\left\{\mathbf{x}:\, \sum_{k=1}^{M} x_k g_k = y\right\}} \varphi(\mathbf{x}). \tag{4.7}$$

In the theory of large deviations of probability, this result is known as contraction principle [38]. $\psi(y)$ and $\varphi(\mathbf{x})$ are called a level-1 and a level-2 large deviations rate functions, respectively.

From phenotypic frequencies to biomarker mean values— The right-hand-side of (4.7) can be further carried out; this is a problem of constrained minimization using multivariate calculus:

$$\begin{cases} \min_{\mathbf{x}} \left\{ \left(\sum_{k=1}^{M} x_k \ln \frac{x_k}{p_k} \right) \right\}, \\ \sum_{k=1}^{M} x_k g_k = y, \\ \sum_{k=1}^{M} x_k = 1. \end{cases} \tag{4.8}$$

Introducing Langrage multipliers for (4.8),

$$\mathscr{L}(\mathbf{x}, \beta, \varphi) = \ln \left(\sum_{k=1}^{M} x_k \ln \frac{x_k}{p_k} \right) + \beta \left(\sum_{k=1}^{M} x_k g_k - y \right) + \lambda \left(\sum_{k=1}^{M} x_k - 1 \right). \tag{4.9}$$

Then we can find $\mathbf{x}^* = (x_1^*, x_2^*, \cdots, x_M^*)$, β^*, and λ^* as the solution of

$$\frac{\partial \mathscr{L}(\mathbf{x}, \beta, \lambda)}{\partial x_j} = \frac{\partial \mathscr{L}(\mathbf{x}, \beta, \lambda)}{\partial \beta} = \frac{\partial \mathscr{L}(\mathbf{x}, \beta, \lambda)}{\partial \lambda} = 0. \tag{4.10}$$

That is,

$$x_k^* = \frac{p_j e^{-\beta^* g_k}}{\sum_{j=1}^{M} p_j e^{-\beta^* g_j}}, \tag{4.11a}$$

$$y = \sum_{j=1}^{M} x_j^* g_j = - \left[\frac{\partial}{\partial \beta} \ln \sum_{j=1}^{M} p_j e^{-\beta g_j} \right]_{\beta=\beta^*}, \tag{4.11b}$$

in which β^* is a function of y through Eq. (4.11b), which gives the function implicitly. We therefore obtain

$$\begin{aligned} \psi(y) &= \varphi(\mathbf{x}^*) = \sum_{k=1}^{M} x_k^* \ln \left(\frac{e^{-\beta^* g_k}}{\sum_{j=1}^{M} p_j e^{-\beta^* g_j}} \right) \\ &= -\beta^* \sum_{k=1}^{M} x_k^* g_k - \ln \sum_{j=1}^{M} p_j e^{-\beta^* g_j} = -\beta^* y + \beta^* F(\beta^*), \end{aligned} \tag{4.12}$$

where

$$F(\beta) = -\frac{1}{\beta} \ln Z(\beta), \ Z(\beta) \equiv \sum_{j=1}^{M} p_j e^{-\beta g_j}, \tag{4.13}$$

and $\beta^*(y)$ solves $\mathrm{d}[\beta F(\beta)]/\mathrm{d}\beta = y$.

The above computation tells us that if one knows the values of a biomarker for all the M states of a cell, $g_1, g_2, \cdots, g_M$, together with a prior knowledge of $p_1, p_2, \cdots, p_M$, one should construct the $Z(\beta)$ function and calculate the $F(\beta)$ given in (4.13). Then the probability distribution for the mean value of the biomarker is going to be:

$$\Pr\left\{\overline{\mathbf{g}}^{(N)} = y\right\} \propto \left[e^{-N\beta\{F(\beta)-y\}}\right]_{\beta=\beta^*(y)}. \tag{4.14}$$

It also tells us that if one observes the mean biomarker value being $\hat{y}$, then the most probable phenotypic frequencies will have a posterior form that deviates from its prior $\{p_k\}$:

$$f_k\Big\{\text{conditioned on } \overline{\mathbf{g}} = \hat{y}\Big\} = \left[\frac{p_k e^{-\beta g_k}}{Z(\beta)}\right]_{\beta=\beta^*(\hat{y})}. \tag{4.15}$$

Both Eqs. (4.14) and (4.15) suggest that the functional relationship $y = \mathrm{d}[\beta F(\beta)]/\mathrm{d}\beta$, between the mean value of the biomarker $y = \overline{\mathbf{g}}$ and the Lagrangian multiplier β, or its inverse form $\beta = \beta^*(y)$, are very fundamental to the probabilistic problem, in the limit of infinite sample size $N \to \infty$.

4.3 Beyond an i.i.d. population

We derived the expression in (4.3) based on the assumption of a population of N cells that are i.i.d. samples of a single M-state random individual with probability $\{p_k\}$. When there are cell-cell interactions among the individuals within a population, the mathematics immediately becomes much more involved.

Two types of research go beyond an i.i.d. population in the stochastic modeling; they were originally motivated, respectively, by chemical kinetics in solution [116] and Ising model for ferromagnetism of solid [125]. In chemical kinetics, rapid spatial movement of all "individual molecules" in an aqueous solution leads to the assumption that every individual collides with every other invidivual, and certain "reactions" can occur randomly. The Gibbs function in chemical thermodynamics is precisely like the $\varphi(\mathbf{x})$ function in Eq. (4.3), for complex chemical reaction systems in equilibrium [74]. Actually there is a general equation, first discovered by G. Hu [68], whose solution can provide $\varphi(\mathbf{x})$ for non-i.i.d. populations. For $J = M - 1$ reversible unimolecular reactions among M species, $X_1, X_2, \cdots, X_M$, with concentrations $\mathbf{x} = (x_1, x_2, \cdots, x_M)$ and arbitrary non-negative functions $R_{\pm j}(\mathbf{x})$ being the rates of the j^{th} reaction between species j and the species M,

$$X_1 \underset{R_{+1}}{\overset{R_{-1}}{\rightleftharpoons}} X_M,\ X_2 \underset{R_{+2}}{\overset{R_{-2}}{\rightleftharpoons}} X_M,\ \cdots,\ X_{M-1} \underset{R_{+J}}{\overset{R_{-J}}{\rightleftharpoons}} X_M, \tag{4.16}$$

the equation reads

$$\sum_{j=1}^{J} R_{+j}(\mathbf{x})\Big[1 - e^{\partial\varphi/\partial x_j - \partial\varphi/\partial x_M}\Big] + R_{-j}(\mathbf{x})\Big[1 - e^{-\partial\varphi/\partial x_j + \partial\varphi/\partial x_M}\Big] = 0. \tag{4.17}$$

If $R_{+j}(\mathbf{x}) = q_j x_M$ and $R_{-j}(\mathbf{x}) = r_j x_j$, then the solution to Eq. 4.17 recovers the (4.3),

$$\varphi(\mathbf{x}) = \sum_{m=1}^{M} x_m \ln\left(\frac{x_m}{p_m}\right), \tag{4.18}$$

in which the p's are functions of q's and r's,

$$p_m = \frac{\frac{q_m}{r_m}}{\frac{q_1}{r_1} + \cdots + \frac{q_{M-1}}{r_{M-1}} + 1}. \tag{4.19}$$

The particular set of $R_{\pm j}(\mathbf{x})$ represents chemical reactions in an ideal solution. A reader who had a course on freshman chemistry might recognize (4.18) as

$$G(\mathbf{x}) = \sum_{m=1}^{M} x_m \mu_m,\ \mu_m(x_m) = \mu_m^o + RT\ln x_m,$$

where μ_m is the chemical potential of m^{th} specie with mole fraction x_m (not molarity) in an ideal solution, and $\mu_m^o = -RT\ln p_m$. Then $\Delta\mu_{ij}^o = \mu_i^o - \mu_j^o = -RT\ln(p_i/p_j)$, where (p_i/p_j) is the equilibrium constant between species i and j [28]. Apart from the RT, the Gibbs energy function is a consequence of statistical counting, which has very little to do with the energy of the atoms in the molecules [164].

In the second type, Ising model and alike, "individual atoms" are located at fixed lattice points in a solid, each one only interacts with its neighbours. The limit of $N \to \infty$ of such an interacting particle system is known as *hydrodynamic limit* of the stochastic model.

Cell-cell interactions in a tissue or in a culture medium can have both types: When an interaction is mediated by rapidly diffusing small molecular factors, one can safely assume the interaction is between every two individual cells in a population. If an interaction between nearby cells is mediated by slowly diffusing molecules, or due to direct contacts via mechanical interactions, gap junctions, or synapses, then a lattice model is more appropriate. Combining these two types of mathematical descriptions leads to the "reaction diffusion" paradigm [141] which serves the foundation for describing living phenomena [205].

4.4 Discussion

Statistical mechanics and Boltzmann's law— ho had a course on statistical mechanics [92] will certainly recognize $Z(\beta)$, $F(\beta)$, and β^{-1} in Eq. (4.13) as partition function, Helmholtz free energy, and temperature, if one identifies g_k as the energy of the k^{th} state of a mechanical system. Eq. (4.12) then shows that $F(\beta) = y + \beta^{-1}\psi(y)$ where $-\psi(y)$ should be identified as "entropy" of the mechanical system with energy y; and it is related to $F(\beta)$ through a Legendre transform. Most textbooks on statistical mechanics do not tell its readers, however, the clear mathematical logic of all these formulae. But actually, Boltzmann's 1877 paper [181], by counting the molecules with different kinetic energy in an ideal gas, had proceeded exactly the steps we took and derived the celebrated Boltzmann's law, in the form in Eq. (4.11a).

Variational Bayesian method — The $F(\beta)$ obtained in Eq. (4.13) has a very important property: For any, arbitrary, *normalized* distribution $\{z_k\}$,

$$\sum_{k=1}^{M} z_k \ln\left(\frac{z_k}{p_k e^{-\beta g_k}}\right) \geq -\ln\left(\sum_{k=1}^{M} p_k e^{-\beta g_k}\right) = \beta F(\beta). \tag{4.20}$$

In the variational Bayesian method for inference [79], one often knows a target, posterior distribution $p_k e^{-\beta g_k}$ but computing its normalization factor is expensive. Eq. (4.20) shows that to obtain the target distribution, one can simply minimizes the left-hand-side of (4.20) among a set of possible $\{z_k\}$. This same idea had also been used by Gibbs in his variational method [148]: The free energy $F(\beta)$ of an equilibrium state is the minumum among all others through a virtual change of state.

Maximum entropy principle— The constrained optimization in (4.8) leading to distribution in (4.11a) has also become the foundation of *maximum entropy principle* (MEP) championed by E.

T. Jaynes [99], which has played a productive role in data science. The axiomatic nature of MEP [183] and the role of conditional probability [196] have been elucidated.

The fundamental premises behind the large deviations principle (LDP) and the MEP are very different: Entropy, as a large deviation rate function, is used in the former to find the rare event that is the most probable, which is the only possible event in the limit: For an arbitrary set of n real values $\{\varphi_i\}$,

$$\left(e^{-N\varphi_1} + e^{-N\varphi_2} + \cdots + e^{-N\varphi_n}\right) \sim e^{-N\varphi^*} \text{ as } N \to \infty, \tag{4.21}$$

where $\varphi^* = \min\{\varphi_1, \cdots, \varphi_n\}$. This is the same idea in choosing only the term with the largest eigenvalue among the terms in a linear eigenvalue decomposition, in the limit of infinite time or system's size. In MEP, however, entropy function is used as a measure for "unbias". Actually, according to LDP, the $\mathbf{x}^*$ in (4.11a) is not a probability distribution, it is the most probable frequency among N i.i.d. samples. In MEP, it is interpreted as the least biased probability distribution with maximum uncertainty.

Part II

ASYMPTOTIC BEHAVIORS OF STOCHASTIC DIFFERENTIAL EQUATIONS

Chapter 5

KINEMATIC BASIS OF EMERGENT ENERGETICS OF COMPLEX DYNAMICS

This chapter is based on joint work with Hong Qian and Ying-Jen Yang [170].

5.1 Introduction

Classical mechanics has traditionally been divided into *kinematics* and *dynamics*. The former gives the precise relationship between a mechanical motion $\mathbf{x}(t)$ and descriptions of the motion in terms of its velocity $\dot{\mathbf{x}}(t)$ and acceleration $\ddot{\mathbf{x}}(t)$ under geometric constraints, and the latter provides relationships between the motions and the concepts of mass, force, and *mechanical energy*. While the former is a part of calculus, the latter constitutes the core of the classical physics of motion. When a mechanical system contains a great many number of point masses such as atoms and molecules, the notions of *heat* and *temperature*, as a stochastic description of complex mechanical motions and kinetic energy, arise.

In classical, macroscopic chemical kinetics, a reaction in aqueous solution, say $A + B \to C$, is described by a rate process: $\mathrm{d}c_A(t)/\mathrm{d}t = -r(t)$, which should be identified as the chemical kinematics. It is again based on calculus which rigirously defines the concept of instantaneous rate of concentration change (fluxion). The functional relationship between $r(t)$ and the concentrations $c_A(t)$ and $c_B(t)$, known as a *rate law*, is analogous to the constitutive relations. One well-known example of a rate law is the Guldberg-Waage mass action $r(t) = kc_A(t)c_B(t)$ where k is a constant independent of c_A and c_B. The notion of energy in chemical reactions, however, exists in the chemical thermodynamics of heterogeneous substants, a separate theory developed by J. W. Gibbs, who introdued the notion of *Gibbs function* and *chemical potential* as the energy and force that drives the chemical changes.

With the above understandings, therefore, it came as a surprise that a recent work [74, 75] claims that the mathematical foundation of Gibbsian chemical thermodynamics, at a given temperature, needs only the mesoscopic stochastic kinematics, irrespective of any details of the mechanics of the atoms and molecules. In other words, the isothermal chemical thermodynamics of Gibbs is dictated purely by the kinematics via the mathematics of probability. The implication of this observation is conceptually sigificant: It implies in general a stochastic description of a complex dynamics has a "hidden" energetics that is already defined by mathematics! The present paper applies this novel idea, *stochastic kinematics dictates energetics*, to another general class of processes and further explore the idea: diffusion processes in continuous space $\mathbb{R}^n$ with continuous time $t \in \mathbb{R}$. Mathematical analysis again reveals a hidden energetic and thermodynamic structure that underlying the kinematics. The term "thermo" here does not imply heat; rather it means a stochastic description of a complex dynamics.

The concept of *thermodynamic force* was clearly articulated in the work of L. Onsager [144]. In chemical kinetics, it is widely accepted that *entropic force* is a legitimate description on par with mechanical force as a "cause" for an action [110, 88]. The Shannon entropy is computable in any statistical description of dynamics [109, 131]. Actually, P. W. Anderson has stated that "[A]t each level of complexity entirely new properties appear, and the understanding of the new behaviors requires research which I think is as fundamental in its nature as any other." [3] In fact, he continued to provide a recipe for discovering an emergent law:

> "It is only as the nucleus is considered to be a many-body system — in what is often called the $N \to \infty$ limit — that such [emergent] behavior is rigorously definable. ... Starting with the fundamental laws and a computer, we would have to do two impossible things — solve a problem with infinitely many bodies [e.g., zero fluctuations], and then apply the result to a finite system — before we synthesized this behavior."

As we shall see, the discovery of the hidden thermodynamic laws indeed involves taking the limit of noise, e.g., fluctuations, tending zero. Our result thus provides a clear understanding of why and how emergent thermodynamic behaviors, as statistical laws, can be independent of the underlying details.

We consider the general description of complex dynamics in terms of a stochastic Markov process $\mathbf{X}_\epsilon(t)$. Such a dynamics can be represented by its probability distribution $p_\epsilon(\mathbf{x},t)$ that follows a Fokker-Planck equation (FPE) [69]

$$\frac{\partial p_\epsilon}{\partial t} = -\nabla\cdot\mathbf{J}\big[p_\epsilon\big],\ \mathbf{J}\big[p_\epsilon\big] \equiv \mathbf{b}(\mathbf{x})p_\epsilon - \epsilon\mathbf{D}(\mathbf{x})\nabla p_\epsilon. \tag{5.1}$$

Its trajectory is represented in terms of the solution to a Langevin type equation $\mathrm{d}\mathbf{X}_\epsilon(t) = \tilde{\mathbf{b}}(\mathbf{X}_\epsilon)\mathrm{d}t + \big[2\epsilon\mathbf{D}(\mathbf{X}_\epsilon)\big]^{\frac{1}{2}}\mathrm{d}\mathbf{B}(t)$, where $\mathbf{B}(t)$ is the standard multidimentional Brownian motion and $\tilde{\mathbf{b}} = \mathbf{b} + \nabla\cdot\epsilon\mathbf{D}$. The ϵ signifies a connection between deterministic and stochastic motions [163]: In the limit of $\epsilon\to 0$, the stochastic trajectory $\mathbf{X}_\epsilon(t)\to\hat{\mathbf{x}}(t)$ which satisties $\mathrm{d}\hat{\mathbf{x}}(t)/\mathrm{d}t = \mathbf{b}\big(\hat{\mathbf{x}}(t)\big)$, and similarly $p_\epsilon(\mathbf{x},t)\to\delta\big(\mathbf{x}-\hat{\mathbf{x}}(t)\big)$ if initial value $p_\epsilon(\mathbf{x},0)=\delta\big(\mathbf{x}-\hat{\mathbf{x}}(0)\big)$.

We emphasize that a stochastic description of dynamics does not have any notion of "energy" and "thermodynamics" at the onset. This is what we meant by "kinematic description". As we shall show, however, in addition to the actual limit $\hat{\mathbf{x}}(t)$, a mathematico-thermodynamic structure also emerges in the process of taking $\epsilon\to 0$: For a fixed ϵ and regarding the stochastic dynamics in Eq. (5.1), it is known that the relative entropy

$$F\big[p_\epsilon(\mathbf{x},t)\big] \equiv \int_{\mathbb{R}^n} p_\epsilon(\mathbf{x},t)\ln\left(\frac{p_\epsilon(\mathbf{x},t)}{\pi_\epsilon(\mathbf{x})}\right)\mathrm{d}\mathbf{x}, \tag{5.2}$$

has a paramount importance [15, 158, 34], where the stationary solution to (5.1) $\pi_\epsilon(\mathbf{x})$ embodies the notion of an entropic force. It was discovered only recently that the F satisfies a balance equation $\mathrm{d}F/\mathrm{d}t \equiv -f_{\mathrm{d}}(t) = Q_{\mathrm{hk}}(t) - e_{\mathrm{p}}(t)$ [70, 197, 55, 161], with

$$\begin{aligned} f_{\mathrm{d}}\big[p_\epsilon\big] &= \int_{\mathbb{R}^n}\mathbf{J}\big[p_\epsilon\big]\cdot\nabla\ln\left(\frac{p_\epsilon}{\pi_\epsilon}\right)\mathrm{d}\mathbf{x}, && (5.3\text{a})\\ Q_{\mathrm{hk}}\big[p_\epsilon\big] &= \int_{\mathbb{R}^n}\mathbf{J}\big[p_\epsilon\big]\cdot(\epsilon\mathbf{D})^{-1}(\mathbf{x})\mathbf{J}\big[\pi_\epsilon\big]\pi_\epsilon^{-1}\mathrm{d}\mathbf{x}, && (5.3\text{b})\\ e_{\mathrm{p}}\big[p_\epsilon\big] &= \int_{\mathbb{R}^n}\mathbf{J}\big[p_\epsilon\big]\cdot(\epsilon\mathbf{D})^{-1}(\mathbf{x})\mathbf{J}\big[p_\epsilon\big]p_\epsilon^{-1}\mathrm{d}\mathbf{x}. && (5.3\text{c})\end{aligned}$$

All three quantities in Eq. (5.3) are non-negative. This fact ensures the interpretation of Q_{hk} and e_{p} as the source and sink for the "free energy" of a system. The $F\big[p_\epsilon\big]$ is interpreted as a generalized, nonequilibrium free energy since $-\epsilon \int_{\mathbb{R}^n} p_\epsilon \ln \pi_\epsilon \mathrm{d}\mathbf{x}$, $-\int_{\mathbb{R}^n} p_\epsilon \ln p_\epsilon \mathrm{d}\mathbf{x}$, and ϵ are analogous to "mean internal energy" E, entropy S, and temperature T, thus $F = E - TS$. See [96, 179, 198] for the relation between these average quantities and the trajectory-based stochastic thermodynamics, Jarzynski like equalities, and fluctuation theorems.

The present work provides a set of new "thermodynamic" relations in the form of a set of three equations in Eqs. (5.12), putting together as a system, and their physical interpretations. As a mathematical result through perturbation theory, the three equations were known, but not with the thermodynamic connections and the geometric meaning shown in present work. Eqs. (5.12a,b) were in the work of Ventsel and Freidlin and the studies that followed [204, 52, 84]. (5.12c) has been derived in the work of Graham and Tél [83]. The latter work focused on constructing asymptotic stationary probability distributions based on the three equations.

The present work illustrates a deep relation between this set of equations as a dynamic description and thermodynamic energetics, in the context of the recently developed stochastic thermodynamics. We regard Eq. (5.12) as an emergent *energetics* that accompanies the *deterministic dynamics* $\dot{\mathbf{x}} = \mathbf{b}(\mathbf{x})$. As a characterization of a complex behavior, the $\mathbf{b}(\mathbf{x})$ does not exist alone as a macroscopic description, it comes with additional information coded in $\mathbf{D}$, φ, and ω, as *thermodynamics*. A new Pythagorean equation (5.20) that relates three entropy productions in (5.19), the macroscopic counterparts of (5.3), is revealed in the present work. This further suggests a geometric perspective to be explored.

We note the Eq. (5.12a) is not a Helmoholtz (or Hodge) decomposition of a vector field since $\boldsymbol{\gamma}(\mathbf{x})$ is not divergence free [162]. The motion following $\boldsymbol{\gamma}(\mathbf{x})$ however, conserves the $\varphi(\mathbf{x})$ according to Eq. (5.12b). The divergence of $\boldsymbol{\gamma}(\mathbf{x})$ is provided by the third equation (5.12c). The emergence of the $\omega(\mathbf{x})$ indicates the significance of a "local volume" of this φ-conservative motion [73]: This is called degeneracy in the classical statistical mechanical termonology. While the motion following $\boldsymbol{\gamma}(\mathbf{x})$ can be complex, a proper statistical description, called physical measure, usually exists [216]. For more discussion of the thermodynamic meanings of dissipative vs. conservative motions, and their relation to trajectory-based entropy production, see [162, 130]. The present work also finds explicit high-order contributions to entropy production, in (5.27), beyond the standard Gaussian fluctuations.

5.2 Emergent potential and φ-conservative motion γ

The mathematical theory of large deviations connects the stochastic dynamics with finite ϵ to the deterministic dynamics with $\epsilon = 0$ [65, 38, 145, 193, 185, 72]. It is a rigorous and complete asymptotic theory akin to the WKB ansatz:

$$p_\epsilon(\mathbf{x}, t) = \exp\big[-\varphi^{\mathrm{td}}(\mathbf{x}, t)/\epsilon + o\big(\epsilon^{-1}\big)\big], \tag{5.4}$$

in which $\varphi^{\mathrm{td}}(\mathbf{x}, t)$ is known as a time-dependent *large deviation rate function*. [1] Taking the Eq. (5.4) as given and recall that $p_\epsilon(\mathbf{x}, t) \to \delta\big(\mathbf{x} - \hat{\mathbf{x}}(t)\big)$, then it is easy to see that in the limit of $\epsilon \to 0$,

$$\epsilon F\big[p_\epsilon(\mathbf{x}, t)\big] \to \varphi^{\mathrm{ss}}\big(\hat{\mathbf{x}}(t)\big),\ \varphi^{\mathrm{ss}}(\mathbf{x}) = -\lim_{\epsilon\to 0} \epsilon \ln \pi_\epsilon(\mathbf{x}). \tag{5.5}$$

Since $\mathrm{d}F/\mathrm{d}t = -f_{\mathrm{d}} \le 0$, $F\big[p_\epsilon(\mathbf{x}, t)\big]$ is a monotonic non-incresing function of t. Consequently, Eq. (5.5) states that $\varphi^{\mathrm{ss}}\big(\hat{\mathbf{x}}(t)\big)$ is a monotonic function of t, which implies that $\varphi^{\mathrm{ss}}(\mathbf{x})$ is an energy function of the dissipative dynamics $\hat{\mathbf{x}}(t)$, the solution to the deterministic equation $\mathrm{d}\mathbf{x}/\mathrm{d}t = \mathbf{b}(\mathbf{x})$. We shall drop the superscript from the steady-state large deviation rate function $\varphi^{\mathrm{ss}}(\mathbf{x})$ from now on.

The connection between $F[p_\epsilon]$ and $\varphi(\mathbf{x})$ illustrates that the latter is an emergent energetic quantity in the limit of $\epsilon \to 0$. It provides a new derivation with thermodynamic insights of the mathematical result of Ventsel and Freidlin [204, 52, 84]. The relation firmly connects (mesoscopic) stochastic free energy $F[p_\epsilon]$ with (macroscopsic) deterministic pseudo-potential $\varphi(\mathbf{x})$; it validates the earlier interpretation of $-\epsilon \int_{\mathbb{R}^n} p_\epsilon \ln \pi_\epsilon \mathrm{d}x$ as a mean internal energy, with $\varphi(\mathbf{x})$ as the internal energy function of state $\mathbf{x} \in \mathbb{R}^n$.

We emphasize that the original $\mathbf{b}(\mathbf{x})$ in Eq. (5.1) is not a gradient field in general. In terms of the newfound $\varphi(\mathbf{x})$ and continue the WKB ansatz, one can express $\pi_\epsilon(\mathbf{x}) = \omega(\mathbf{x}) \exp[-\varphi(\mathbf{x})/\epsilon + a \ln \epsilon + O(\epsilon)]$, in which $\ln \omega(\mathbf{x})$ is the next order to the leading two terms and $a \ln \epsilon$ arising from

[1] The $\varphi^{\mathrm{td}}(\mathbf{x}, t)$ satisfies a time-dependent equation of its own: $\partial\varphi^{\mathrm{td}}(\mathbf{x}, t)/\partial t = -\nabla\varphi^{\mathrm{td}}(\mathbf{x}, t) \cdot \boldsymbol{\gamma}^{\mathrm{td}}(\mathbf{x}, t)$, where $\boldsymbol{\gamma}^{\mathrm{td}}(\mathbf{x}, t) \equiv \mathbf{D}(\mathbf{x})\nabla\varphi^{\mathrm{td}}(\mathbf{x}, t) + \mathbf{b}(\mathbf{x})$. This is known as a Hamilton-Jacobi equation.

normalization fractor is independent of $\mathbf{x}$. Substituting this expression into the stationary FPE

$$-\nabla \cdot \mathbf{J}[\pi_\epsilon(\mathbf{x})] = \nabla \cdot \big(\epsilon \mathbf{D}(\mathbf{x})\nabla \pi_\epsilon(\mathbf{x}) - \mathbf{b}(\mathbf{x})\pi_\epsilon(\mathbf{x})\big) = 0, \tag{5.6}$$

one obtains

$$\begin{aligned}\epsilon^{-1}\omega(\mathbf{x})\boldsymbol{\gamma}(\mathbf{x}) \cdot \nabla\varphi - \big[\nabla\omega(\mathbf{x}) \cdot \mathbf{D}(\mathbf{x})\nabla\varphi + \nabla \cdot \big(\omega(\mathbf{x})\boldsymbol{\gamma}(\mathbf{x})\big)\big]& \\ +\epsilon\nabla \cdot \Big(\mathbf{D}(\mathbf{x})\nabla\omega(\mathbf{x})\Big) + O(\epsilon) = 0,&\end{aligned} \tag{5.7}$$

in which $\boldsymbol{\gamma}(\mathbf{x}) \equiv \mathbf{D}(\mathbf{x})\nabla\varphi(\mathbf{x}) + \mathbf{b}(\mathbf{x})$. Equating like order terms in Eq. (5.7), we have

$$\nabla\varphi(\mathbf{x}) \cdot \boldsymbol{\gamma}(\mathbf{x}) = 0,\ \forall \mathbf{x} \in \mathbb{R}^n, \tag{5.8}$$

the vector field $\boldsymbol{\gamma}(\mathbf{x})$ is orthogonal to $\nabla\varphi(\mathbf{x})$, and [83]

$$\nabla \cdot \big(\omega(\mathbf{x})\boldsymbol{\gamma}(\mathbf{x})\big) = -\nabla\omega(\mathbf{x}) \cdot \mathbf{D}(\mathbf{x})\nabla\varphi(\mathbf{x}). \tag{5.9}$$

Actually, $\boldsymbol{\gamma}_\epsilon(\mathbf{x}) \equiv \pi_\epsilon^{-1}(\mathbf{x})\mathbf{J}[\pi_\epsilon(\mathbf{x})]$ has been identified as Onsager's thermodynamic flux[157], and γ is its limit as $\epsilon \to 0$:

$$\lim_{\epsilon\to 0} \boldsymbol{\gamma}_\epsilon(\mathbf{x}) = \mathbf{b}(\mathbf{x}) + \mathbf{D}(\mathbf{x})\nabla\varphi(\mathbf{x}) \equiv \boldsymbol{\gamma}(\mathbf{x}). \tag{5.10}$$

The motions following the vector field $\boldsymbol{\gamma}(\mathbf{x})$ is restricted on the level set of $\varphi(\mathbf{x})$. This result also has a correspondence when ϵ is finite [82, 173]: The stationary FPE can be re-written as

$$\nabla\varphi_\epsilon(\mathbf{x}) \cdot \boldsymbol{\gamma}_\epsilon(\mathbf{x}) = \epsilon\nabla \cdot \boldsymbol{\gamma}_\epsilon(\mathbf{x}), \tag{5.11}$$

in which $\varphi_\epsilon(\mathbf{x}) \equiv -\epsilon \ln \pi_\epsilon(\mathbf{x})$ has been widely called a kinetic potential [82, 112, 143]. Therefore, if $\nabla \cdot \boldsymbol{\gamma}_\epsilon = 0\ \forall\epsilon$, then there is an orthogonality between $\nabla\varphi_\epsilon(\mathbf{x})$ and $\boldsymbol{\gamma}_\epsilon(\mathbf{x})$ for all ϵ. More generally, irrespective of $\nabla \cdot \boldsymbol{\gamma}_\epsilon$ being zero or not, in the limit of $\epsilon \to 0$, $\varphi_\epsilon(\mathbf{x}) \to \varphi(\mathbf{x})$, $\boldsymbol{\gamma}_\epsilon \to \boldsymbol{\gamma}$, and $\boldsymbol{\gamma} \cdot \nabla\varphi = 0$.

$\boldsymbol{\gamma}(\mathbf{x}) = 0$ is mathematically equivalent to detailed balance. Stochastic systems with $\boldsymbol{\gamma} = 0$ is widely considered as "non-driven" [220, 76] and is expected to approach to an equilibrium steady state in the long-time limit. For such systems, the free energy F aquires additional meaning as the potential of Onsager's thermodynamics force, $\boldsymbol{\gamma}_\epsilon = \mathbf{D}(\mathbf{x})\nabla F$.

Collecting Eqs. (5.8), (5.9) and (5.10), we have a system of three equations

$$\mathbf{b}(\mathbf{x}) = -\mathbf{D}(\mathbf{x})\nabla\varphi(\mathbf{x}) + \boldsymbol{\gamma}(\mathbf{x}), \tag{5.12a}$$

$$\nabla\varphi(\mathbf{x})\cdot\boldsymbol{\gamma}(\mathbf{x}) = 0, \tag{5.12b}$$

$$\nabla\cdot\big(\omega(\mathbf{x})\boldsymbol{\gamma}(\mathbf{x})\big) = -\nabla\omega(\mathbf{x})\cdot\mathbf{D}(\mathbf{x})\nabla\varphi(\mathbf{x}), \tag{5.12c}$$

in which the vector fields $\mathbf{b}(\mathbf{x})$ and $\boldsymbol{\gamma}(\mathbf{x})$ represent dynamics, $\mathbf{D}(\mathbf{x})$, which represents stochastic motion, can be thought as a geometric metrics, $\omega(\mathbf{x})$ represents local "measure" for the state space volume (degeneracy in the statistical mechanical termonology)[2], $\varphi(\mathbf{x})$ and $\ln\omega(\mathbf{x})$ are thermodynamic quantities akin to energy and entropy, respectively. The "noise structure" $\mathbf{D}(\mathbf{x})$ provides a unique geometry for the dynamics.

In the simplest case, if $\mathbf{D}(\mathbf{x})$ is the identity matrix, and $\omega(\mathbf{x}) = 1$ is the Lebesgue measure, then the equations in Eq. (5.12) become

$$\mathbf{b}(\mathbf{x}) = -\nabla\varphi(\mathbf{x}) + \boldsymbol{\gamma}(\mathbf{x}), \tag{5.13a}$$

$$\boldsymbol{\gamma}(\mathbf{x})\cdot\nabla\varphi(\mathbf{x}) = 0, \tag{5.13b}$$

$$\nabla\cdot\boldsymbol{\gamma}(\mathbf{x}) = 0. \tag{5.13c}$$

The vector field $\boldsymbol{\gamma}(\mathbf{x})$ is now volume preserving, and the system $\mathbf{x}'(t) = \gamma(\mathbf{x})$ also has a conserved quantity $\varphi(\mathbf{x})$. System with Eq. (5.13) is intimately related to the classical Hamiltonian systems [151]. One of the most important features of this class of dynamics is that the $\varphi(\mathbf{x})$ gives the steady state probability distribution exactly for any finite ϵ in the form of $\pi_\epsilon(\mathbf{x}) \propto e^{-\varphi(\mathbf{x})/\epsilon}$ according to Boltzmann's law, if one identifies ϵ with temperature [162, 163].

[2] Actually, there is also a time-dependent equation for $\omega^{\mathrm{td}}(\mathbf{x},t)$: $\partial\omega^{\mathrm{td}}(\mathbf{x},t)/\partial t = -\big(\mathbf{D}(\mathbf{x})\nabla\varphi^{\mathrm{td}}(\mathbf{x},t)\big)\cdot\nabla\omega^{\mathrm{td}}(\mathbf{x},t) - \nabla\cdot\big(\gamma^{\mathrm{td}}(\mathbf{x},t)\omega^{\mathrm{td}}(\mathbf{x},t)\big)$. The asymptotic expansion to this next-order also provides a natural viscosity solution for the HJE with a diffusion term $\epsilon\nabla\cdot\big(\omega^{\mathrm{td}}(\mathbf{x},t)\mathbf{D}(\mathbf{x})\nabla\varphi^{\mathrm{td}}(\mathbf{x},t)\big)$.

5.3 φ-based statistical mechanics and ensemble change

It is seen immediately that if one computes a partition function from energy function $\varphi(\mathbf{x})$ and degeneracy $\omega(\mathbf{x})$:

$$Z(\epsilon) = \int_{\mathbb{R}^n} \omega(\mathbf{x}) e^{-\varphi(\mathbf{x})/\epsilon} \mathrm{d}\mathbf{x}, \tag{5.14}$$

then $Z^{-1}(\epsilon)\omega(\mathbf{x})e^{-\varphi(\mathbf{x})/\epsilon}$ is the asymptotic probability density for $\pi_\epsilon(\mathbf{x})$. If the principle of equal probability is valid, e.g., $\nabla \cdot \boldsymbol{\gamma}(\mathbf{x}) = 0$, then it is the stationary probability density of Eq. (5.1) for all $\epsilon > 0$.

Let us now consider a bivariate stochastic dynamics with $\mathbf{x}$ and y which is assumed to be a scalar for simplicity. The stationary joint probability for $\mathbf{x}$ and y, $p_\epsilon(\mathbf{x}, y)$ is related to the stationary conditional probability $p_\epsilon(\mathbf{x}|y)$ through the marginal distribution for variable y, $p_{\epsilon,y}(y)$: $p_\epsilon(\mathbf{x}, y) = p_\epsilon(\mathbf{x}|y)p_{\epsilon,y}(y)$. In the asmptotic limit of $\epsilon \to 0$, this yields

$$\varphi(\mathbf{x}, y) = \varphi(\mathbf{x}|y) + \varphi_y(y), \tag{5.15}$$

and the partition functions

$$Z_{\mathbf{x},y}(\epsilon) = \int_y Z_{\mathbf{x}|y}(\epsilon; y) e^{-\varphi_y(y)/\epsilon} \mathrm{d}y. \tag{5.16}$$

The $Z_{\mathbf{x}|y}$ is the partition function with fixed y, treated as a parameter, and $Z_{\mathbf{x},y}$ is the partition function with fluctuating y. To asymptotically evaluating the integral in Eq. (5.16), it can be shown that at $y = \overline{y}$, the mean value of the fluctuating y:

$$\frac{\partial}{\partial y}\varphi_y(\overline{y}) = \epsilon \left[\frac{\partial}{\partial y} \ln Z_{\mathbf{x}|y}(\epsilon; y) \right]_{y=\overline{y}} \equiv \xi_y, \tag{5.17}$$

where ξ_y is the conjugate variable to y. Therefore, the integrand in Eq. (5.16) can be approximately expressed as

$$\varphi_y(y) \simeq \varphi_y(\overline{y}) + \xi_y\big(y - \overline{y}\big). \tag{5.18}$$

Eqs. (5.16), (5.17), and (5.18) constitute J. W. Gibbs' theory of ensemble change. In doing so, the thermodynamics of a stationary system with fluctuating y and the thermodynamics of a stationary system with fixed y are logically connected via the large deviation theory. This derivation shares the same spirit as Helmholtz and Boltzmann's 1884 mechanical theory of heat [67, 25]: Both extend the notion of energy from a system with a fixed "parameter" y to an entire family of systems with different y's [129].

5.4 An instantaneous deterministic energy balance equation

While the $\epsilon F\big[p_\epsilon(\mathbf{x},t)\big] \to \varphi\big(\hat{\mathbf{x}}(t)\big)$ as $\epsilon \to 0$ and $p_\epsilon(\mathbf{x},t) \to \delta\big(\mathbf{x}-\hat{\mathbf{x}}(t)\big)$, the free energy dissipation, house-keeping heat, and entropy production rates [70, 161], also known as non-adiabatic, adiabatic, and total entropy production rates [55], become

$$\epsilon f_{\mathrm{d}}\big[p_\epsilon(\mathbf{x},t)\big] \;\to\; \Big[\nabla\varphi(\mathbf{x})\cdot\mathbf{D}(\mathbf{x})\nabla\varphi(\mathbf{x})\Big]_{\mathbf{x}=\hat{\mathbf{x}}(t)}, \tag{5.19a}$$

$$\epsilon Q_{\mathrm{hk}}\big[p_\epsilon(\mathbf{x},t)\big] \;\to\; \Big[\boldsymbol{\gamma}(\mathbf{x})\cdot\mathbf{D}^{-1}(\mathbf{x})\boldsymbol{\gamma}(\mathbf{x})\Big]_{\mathbf{x}=\hat{\mathbf{x}}(t)}, \tag{5.19b}$$

$$\epsilon e_{\mathrm{p}}\big[p_\epsilon(\mathbf{x},t)\big] \;\to\; \Big[\mathbf{b}(\mathbf{x})\cdot\mathbf{D}^{-1}(\mathbf{x})\mathbf{b}(\mathbf{x})\Big]_{\mathbf{x}=\hat{\mathbf{x}}(t)}. \tag{5.19c}$$

All three quantities are non-negative. They are linked through a Pythagorean-like equation, $\forall\mathbf{x}\in\mathbb{R}^n$:

$$\big\|\mathbf{D}(\mathbf{x})\nabla\varphi(\mathbf{x})\big\|^2+\big\|\boldsymbol{\gamma}(\mathbf{x})\big\|^2=\big\|\mathbf{b}(\mathbf{x})\big\|^2, \tag{5.20}$$

under the inner product $\langle\mathbf{u},\mathbf{v}\rangle\equiv\mathbf{v}\mathbf{D}^{-1}\mathbf{u}$ and thus $\|\mathbf{u}\|^2=\mathbf{u}\mathbf{D}^{-1}\mathbf{u}$.[3]

In the zero-noise limit, the deterministic motion follows $\mathrm{d}\mathbf{x}(t)/\mathrm{d}t=\mathbf{b}(\mathbf{x})$; the balanace equation $\mathrm{d}F/\mathrm{d}t\equiv -f_{\mathrm{d}}=Q_{\mathrm{hk}}-e_{\mathrm{p}}$ now becomes

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\varphi\big(\mathbf{x}(t)\big) &= \mathbf{b}(\mathbf{x})\cdot\nabla\varphi(\mathbf{x})\\ &= \underbrace{\boldsymbol{\gamma}(\mathbf{x})\cdot\mathbf{D}^{-1}(\mathbf{x})\boldsymbol{\gamma}(\mathbf{x})}_{\text{non-conservative pump}}-\underbrace{\mathbf{b}(\mathbf{x})\cdot\mathbf{D}^{-1}(\mathbf{x})\mathbf{b}(\mathbf{x})}_{\text{energy dissipation}}.\end{aligned} \tag{5.21}$$

Both terms before and after the minus sign in Eq. (5.21) are non-negative. Eq. (5.21) constitutes a deterministic, instantaneous energy balance law with the non-conservative pump as the source and

[3] This Pythagorean relation actually exists for stochastic $f_{\mathrm{d}}, Q_{\mathrm{hk}}, e_{\mathrm{p}}$ with finite ϵ if one defines inner product

$$\langle\mathbf{u}(\mathbf{x}),\mathbf{v}(\mathbf{x})\rangle\equiv\epsilon^{-1}\int_{\mathbb{R}^n}\mathbf{u}(\mathbf{x})\mathbf{D}^{-1}(\mathbf{x})\mathbf{v}(\mathbf{x})p_\epsilon(\mathbf{x},t)\mathrm{d}\mathbf{x},$$

and identifies $f_{\mathrm{d}}=\|\epsilon\mathbf{D}(\mathbf{x})\nabla\ln(p_\epsilon(\mathbf{x},t)/\pi_\epsilon(\mathbf{x}))\|^2$, $Q_{\mathrm{hk}}=\|\mathbf{b}(\mathbf{x})-\epsilon\mathbf{D}(\mathbf{x})\nabla\ln\pi_\epsilon(\mathbf{x})\|^2$, and $e_{\mathrm{p}}=\|\mathbf{b}(\mathbf{x})-\epsilon\mathbf{D}(\mathbf{x})\nabla\ln p_\epsilon(\mathbf{x},t)\|^2$.

energy dissipation as the sink, respectively. This result provides a rigorous notion of "energy" for complex dynamics with a stochastic kinematic description. For a classical mechanical system with potential force and friction, φ is the sum of kinetic energy and potential energy, and the energy dissipation is due to the friction.

5.5 Perturbation theory of random processes and the higher-order approximation of e_{p}

From Eq. (5.19), we have shown that the entropy production rate e_{p} is "extensive", i.e., $O\big(\epsilon^{-1}\big)$, and derived it's leading order as $\epsilon \to 0$. This leading order constitutes the part of entropy produced from the deterministic path of the "dissipative dynamics" $\hat{\mathbf{x}}'(t) = \mathbf{b}(\hat{\mathbf{x}})$. We further obtain a result of higher-order terms of the entropy production rate, which yields the part of entropy production around the deterministic trajectory due to the infinitesimal fluctuation. We can visualize this by imaging that we follow the deterministic trajectory and measure the entropy production with a zoomed-in scale. J. Keizer first realized this perspective provides a flucutation-dissipation theorem beyond equilibrium [106, 107].

From a mathematical standpoint, the dissipation part of entropy production is on the scale of the law of large number, while the fluctuation part is on the scale for the central limit theorem. This suggests us that the latter can be obtained by the perturbation of random processes with an appropriate scale. Introducing $\mathbf{Z}_\epsilon(t) = \frac{1}{\sqrt{\epsilon}}(\mathbf{X}_\epsilon(t) - \hat{\mathbf{x}}(t))$ and following the usual perturbation approach [65] with expansion $\mathbf{Z}_\epsilon = \mathbf{Z}^{(0)} + \sqrt{\epsilon}\mathbf{Z}^{(1)} + \cdots + \sqrt{\epsilon}^k\mathbf{Z}^{(k)} + \cdots$, we can write down a stochastic differential equation (SDE) for the $\mathbf{Z}^{(0)}(t)$

$$\mathrm{d}\mathbf{Z}^{(0)}(t) = \mathbf{A}(\hat{\mathbf{x}}(t))\mathbf{Z}^{(0)}(t)\mathrm{d}t + [2\mathbf{D}(\hat{\mathbf{x}}(t))]^{\frac{1}{2}}\mathrm{d}\mathbf{B}(t), \tag{5.22}$$

where $\mathbf{A}(\mathbf{x})$ is the Jacobian matrix of $\mathbf{b}(\mathbf{x})$. Eq. (5.22) is a time-inhomogeneous linear SDE which can be solved as $\mathbf{Z}^{(0)}(t) \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, where $\mathcal{N}$ represents Gaussian distribution and the covariance matrix $\boldsymbol{\Sigma}$ satisfies the equation

$$\frac{\mathrm{d}\boldsymbol{\Sigma}(t)}{\mathrm{d}t} = \mathbf{A}(\hat{\mathbf{x}}(t))\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{A}(\hat{\mathbf{x}}(t))^T + 2\mathbf{D}(\hat{\mathbf{x}}(t)). \tag{5.23}$$

This equation under more restricted consideration had been obtained in [69, 107]. In addition to the small noise expansion for the SDE, we can also expand the FPE of the scaled process $\mathbf{Z}_\epsilon$:

$$\hat{p}_\epsilon(\mathbf{z},t) = \sum_{n=0}^{\infty} \hat{p}_n(\mathbf{z},t)\sqrt{\epsilon}^n, \tag{5.24}$$

in which the $\hat{p}_0(\mathbf{z},t)$ corresponds exactly to the probability distribution of $\mathbf{Z}^{(0)}(t) \sim \mathcal{N}(0,\Sigma)$ [69]. By the change of variable $p_\epsilon(\mathbf{x},t) = \frac{1}{\sqrt{\epsilon}}\hat{p}_\epsilon(\mathbf{z},t)$ and plugging it into the equation of entropy production rate in Eq. (5.3), we can obtain the higher-order approximation

$$\begin{aligned} \epsilon e_{\mathrm{p}}[p_\epsilon] &= \big[\mathbf{b}(\mathbf{x})\cdot\mathbf{D}^{-1}(\mathbf{x})\mathbf{b}(\mathbf{x})\big]_{\mathbf{x}=\hat{\mathbf{x}}(t)} \\ &+ \epsilon\big[\mathrm{tr}\big(\mathbf{M}(\mathbf{x})\big) + 2\mathbf{b}(\mathbf{x})\cdot\mathbf{D}(\mathbf{x})^{-1}\mathbf{m}\big]_{\mathbf{x}=\hat{\mathbf{x}}(t)} + o(\epsilon), \end{aligned} \tag{5.25}$$

where

$$\begin{aligned} \mathbf{M}(\mathbf{x}) &= \mathbf{D}(\mathbf{x})\mathbf{\Sigma}^{-1} + 2\mathbf{A}(\mathbf{x}) + \mathbf{A}(\mathbf{x})^T\mathbf{D}(\mathbf{x})^{-1}\mathbf{A}(\mathbf{x})\mathbf{\Sigma} \\ &+ \mathbf{b}(\mathbf{x})\cdot\mathbf{D}(\mathbf{x})^{-1}\mathbf{H}(\mathbf{x})\mathbf{\Sigma} \end{aligned} \tag{5.26}$$

Note that $\mathbf{H}(\mathbf{x})$ is a rank 3 tensor, for the vector $\mathbf{v} = \mathbf{b}(\mathbf{x})\cdot\mathbf{D}(\mathbf{x})^{-1}$, $\mathbf{vH}(\mathbf{x}) = \sum_i v_i\mathbf{H}_i(\mathbf{x})$, in which $\mathbf{H}_i(\mathbf{x})$ is the Hessian matrix of $b_i(\mathbf{x})$; $\mathbf{\Sigma}$ follows Eq. (5.23), and $\mathbf{m} = \int \mathbf{z}\hat{p}_1(\mathbf{z},t)\mathrm{d}\mathbf{z}$. Therefore, the ϵ order of the entropy production has the form

$$\underbrace{\mathrm{tr}\big(\mathbf{M}(\hat{\mathbf{x}}(t))\big)}_{\text{Gaussian fluctuation}} + \underbrace{2\mathbf{b}(\hat{\mathbf{x}}(t))\cdot\mathbf{D}(\hat{\mathbf{x}}(t))^{-1}\mathbf{m}}_{\text{non-Gaussian fluctuation}}, \tag{5.27}$$

where the first part involves $\mathbf{\Sigma}$ (the second moment with respect to the Gaussian distribution $\hat{p}_0(\mathbf{z},t)$) and the second part involves $\mathbf{m}$ (the first moment with respect to the next order $\hat{p}_1(\mathbf{z},t)$ in the expansion (5.24)). Interestingly, if $\mathbf{Z}_\epsilon(t)$ is a time-inhomogeneous Ornstein-Uhlenbeck process, then the non-Gaussian fluctuation part is always zero. On the other hand, since the non-Gaussian part is due to the nonlinearity of the vector filed $\mathbf{b}$, this part of entropy production rate in the higher order exists uniquely in nonlinear dynamics, e.g. in limit cycles.

5.6 Discussion

In the theory of ordinary differential equations, Hamiltonian dynamics with the conserved H functions and gradient systems with potential functions are two special classes that have been extensively studied [151]. For a general nonlinear dynamics $\dot{\mathbf{x}} = \mathbf{b}(\mathbf{x})$, it is not known whether it always has an associated "energetics". Ventsel and Freidlin's large deviation theory revealed that if $\mathbf{b}(\mathbf{x})$ is the zero-noise limit of a noisy dynamics, a global quasi-potential function $\varphi(\mathbf{x})$ exists [204, 52, 82]. In fact, $\mathbf{b}(\mathbf{x}) = \mathbf{D}(\mathbf{x})\nabla\varphi(\mathbf{x}) + \boldsymbol{\gamma}(\mathbf{x})$, where $\boldsymbol{\gamma}(\mathbf{x}) \perp \nabla\varphi(\mathbf{x})$ at every $\mathbf{x}$. Graham and Tél further showed [83] $\nabla \cdot \big(\omega(\mathbf{x})\boldsymbol{\gamma}(\mathbf{x})\big) = -\nabla\omega(\mathbf{x}) \cdot \mathbf{D}(\mathbf{x})\nabla\varphi(\mathbf{x})$, where $\omega(\mathbf{x})$ represents a proper local measure at $\mathbf{x}$. The present work shows that this system of dynamic equations has a meaningful thermodynamic interpretation via a geometric equation: $\|\mathbf{b}\|^2 = \|\mathbf{D}\nabla\varphi\|^2 + \|\boldsymbol{\gamma}\|^2$, corresponding to the instantaneous rates of total entropy production, free energy dissipation, and house-keep heat, respectively.

If one identifies $\dot{\mathbf{x}} = \mathbf{b}(\mathbf{x})$ as a kinematic description of a complex dynamics, then the system in Eq. (5.12) and Eq. (5.21) provide an energetic description that is hidden under the kinematics. A few words concerning the role of $\mathbf{D}(\mathbf{x})$ are in order. The concept of a gradient field on $\mathbb{R}^n$ requires a notion of *distance*. This is naturally provided by the noise structure embedded in $\mathbf{D}(\mathbf{x})$. This is precisely A. N. Kolmogorov's insights on the nature of probability theory: One needs to have a probability given before carrying out probabilistic computations.

For complex systems, not all "stochasticity" are due to thermal noises. In fact, the Mori-Zwanzig theory of projection operator clearly shows that [152] the dynamics of a projection necessarily induces, in general, a stochastic term and a non-Markovian memory term.

When $\boldsymbol{\gamma} = 0$, the FPE in Eq. (5.1) is a gradient flow in a proper mathematical space [137]. In terms of the Pythagorean-like equation in Eq. (5.20), the stochastic dynamics following the Eq. (5.1) with $\boldsymbol{\gamma} = 0$ has a maximal f_{d}: One leg of a "triangle" has the same length as the triangle's hypotenuse. Thus, under an appropriate geometry, Onsager's *principle of maximum dissipation* can be generalized to nonlinear regime [138]. In recent years, the theory of nonequilibrium landscape has gained wider recognitions in biological physics [5, 172]. Since for a nonequilibrium stochastic system, its stationary state still has highly complex motions as NESS flux [220, 76, 207], the φ is only half the story: the $\boldsymbol{\gamma}$ and its related ω provide the characterization of the NESS motion on each and every φ level set.

Chapter 6

STOCHASTIC LIMIT-CYCLE OSCILLATIONS OF A NONLINEAR SYSTEM UNDER RANDOM PERTURBATIONS

This chapter is based on joint work with Hong Qian [30].

6.1 Introduction

Newtonian mechanics represents the world in terms of featureless point masses with their positions and momenta. In contrast, classical *chemical kinetics* represents the world in terms of the number densities of interacting populations of individual molecules, each with a large internal degrees of freedom, as chemical species. What is possibly an appropriate representation for complex biological systems and processes? To answer this question, it is necessary to to give a more precise meaning to the too widely used term "complex" [90]. Let us consider one class of complex systems, the living biological cells in terms of a biochemical kinetic description. In this paradigm, a *complex system* consists of many interacting sub-populations of individuals with stochastic state transitions; the system as a whole actively exchanges matters, energy, or information with its environment [149]. One sees a remarkable resemblance between this kinetic description of cells and many other biological systems with complex "individuals". In fact, the biochemist's perspective captures a repeated hierarchical structure of the complex world: An ecological system is a community of various biological organism; a human body consists of over 30 trillion cells; and a cell involves a large number of interacting non-living biopolymers. This view echos the philosophy of P. W. Anderson's hierarchical structure of science [3].

While the "stochasticity" in chemical kinetics mainly originates from internal states of individual macromolecules, uncertainties in mechanical motions in biology, such as protein motor proteins in axonal transport and hemodynamics of cardiovascular systems, are chiefly a consequence of *coarse graining*: A highly complex many-body systems can be represented by simple statistical laws. One of the best examples of this is Kramers' rate theory for barrier crossing between two basins, which condenses a very complex dynamics into a simple exponentially distributed time with a single parameter. A problem becomes simple if we focus on the emergent behavior of an assembly of a large numbers of atoms at the macroscopic scale with a much longer time scale. Indeed, experimentalists would find that the macromolecular movement obeys simple laws under certain approximations, for example Fick's law. The bridge between complexity and simplicity is *uncertainty and its statistics*. This is the fundamental idea of the theory of Brownian motion [14].

6.1.1 Stochastic models of complex systems

As can be seen from the above discussion, both representations have their own values for complex systems. Once we choose one of them to describe a system of interest, then the following question is what mathematical model we should adopt. In stochastic chemical kinetics, there is a success of the well-established scaling hypothesis in the continuous-time non-negative integer valued Markov jump process [116]. Consider a continuous stirred chemical reaction vessel of volume V, in which the numbers of molecules of various species $\mathbf{n}_V(t)$ is a Markov jump process that can be described by a *master equation*

$$\frac{\partial P(\mathbf{n}_V,t)}{\partial t} = \sum_{\mathbf{r}} \left[W(\mathbf{n}_V-\mathbf{r},\mathbf{r})P(\mathbf{n}_V-\mathbf{r},t) - W(\mathbf{n}_V,\mathbf{r})P(\mathbf{n}_V,t)\right], \tag{6.1}$$

where $W(\mathbf{n}_V,\mathbf{r})$ is the transition probability per unit time from $\mathbf{n}_V, \mathbf{n}_V+\mathbf{r}$, and both $\mathbf{n}_V$ and $\mathbf{r}$ are q-dimensional vectors. As the system's size $V \to \infty$, $\mathbf{n}_V(t)$ follows the law of large number, $V^{-1}\mathbf{n}_V(t) \to \mathbf{c}(t)$, the concentration of q species.

With a proper scaling by the size V and the assumption that W and P are smooth enough functions, we can take the Kramers-Moyal expansion of the master equation (6.1) [111, 140, 91]

$$\epsilon\frac{\partial p(\mathbf{x},t)}{\partial t} = \sum_{\mathbf{k}} \left(\frac{1}{\mathbf{k}!}\right)\left(\epsilon\frac{\partial}{\partial \mathbf{x}}\right)^{\mathbf{k}} \left[\alpha_{\mathbf{k}}(\mathbf{x})p(\mathbf{x},t)\right], \tag{6.2}$$

where $\epsilon = 1/V$, $\mathbf{x} = \mathbf{n}_V/V$, $p(\mathbf{x},t) = VP(\mathbf{n}_V,t)$, and $\mathbf{k} = (k_1,k_2,\cdots,k_q)$, $\sum_{\mathbf{k}} = \sum_{k_q}\cdots\sum_{k_2}\sum_{k_1}$, $\mathbf{k}! = \prod_i(k_i!)$, and $\alpha_{\mathbf{k}}(\mathbf{x}) = \sum_{\mathbf{r}}(\prod_i r_i^{k_i})w(\mathbf{x},\mathbf{r})$, $w(\mathbf{x},\mathbf{r}) = W(\mathbf{X},\mathbf{r})/V$. The solutions of the differential equation (6.2) with the infinite terms represents the exact time-dependent probability density of the scaled number of population $\mathbf{n}_V/V$. Then a natural question arises: could we obtain a corresponding diffusion process from this infinite order differential equation? The truth is that we can only get a "local diffusion process approximation" for the scaled Markov jump process due to the following reason.

A common method to attack Eq. (6.2) is by truncating the higher order terms to the second order of ϵ to obtain a Fokker-Planck equation (FPE). However, van Kampen [199] pointed out that this method may fail if $\mathbf{n}_V(t)$ has large sizes of jumps. A concrete example was provided in the work [202]: There exists an inconsistency between the stationary solutions of Kramers-Moyal FPE and the original master equation for the Schlögl's model of a chemical reaction system which has bistable steady states. The reason for the failure of Kramers-Moyal FPE is that we are only able to observe either the deterministic behavior of the process at the scale of the law of large number or the Gaussian fluctuations at the scale of the central limit theorem; however, there is no one scale to obtain both. To keep the first two order terms of the Kramers-Moyal expansion simultaneously to represent a diffusion process at a single scale is incorrect. Therefore, van Kampen [199] suggested the Ω expansion which allows us to get a deterministic trajectory as $\epsilon \to 0$ and a local approximation near the deterministic trajectory at the scale $O(\sqrt{\epsilon})$ separately.

In the present work, we focus on the continuous representation of complex systems. We always start with random perturbations of dynamical systems represented by a sequence of stochastic differential equations (SDEs) parameterized by a small parameter ϵ

$$\mathrm{d}\mathbf{X}_\epsilon(t) = \mathbf{b}(\mathbf{X}_\epsilon)\mathrm{d}t + [2\epsilon\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}(t), \tag{6.3}$$

where $\mathbf{X}_\epsilon \in \mathbb{R}^n$, $\mathbf{b} : \mathbb{R}^n \to \mathbb{R}^n$ stands for a drift function, the $\mathbb{R}^n \times \mathbb{R}^n$ diffusion matrix $\mathbf{D}$ is constant and positive semidefinite symmetric, and $\mathbf{B}$ is the standard n-dimensional Brownian motion. This Langevin type equation is widely applicable for complex systems related to mechanics, and it gives us a clear picture of the entire dynamics including the drift and diffusion at one scale. Furthermore, by the rigorous mathematical theory of semigroup [58], every diffusion process represented by a SDE has a unique FPE to characterize the corresponding transition probability density $p_\epsilon(\mathbf{x}, t)$

$$\frac{\partial p_\epsilon}{\partial t} = -\nabla \cdot \mathbf{J}[p_\epsilon], \quad \mathbf{J}[p_\epsilon] \equiv \mathbf{b}(\mathbf{x})p_\epsilon - \epsilon \mathbf{D}\nabla p_\epsilon. \tag{6.4}$$

This line of reasoning to relate diffusion processes and FPEs has no ambiguity unlike the Kramers-Moyal FPEs.

6.1.2 Random perturbations of diffusion processes

Our analysis of random perturbations of diffusion processes is by expansion in powers of ϵ for the sequence of SDEs (6.3), which follows the work of Freidlin and Wentzell [65]. As $\epsilon \to 0$, the sequence of SDEs converges to an ordinary differential equation (ODE) of the emergent deterministic trajectory by the Law of large number (LLN). To shift the sequence of SDEs to its deterministic trajectory with normalization by the scale $O(\sqrt{\epsilon})$, the rescaled sequence of SDEs converges to a time-inhomogeneous Gaussian process by the central limit theorem (CLT). For rare events in $O(1)$, they have the probability asymptotic to zero exponentially fast by the Large deviation principe (LDP). In comparison to Freidlin and Wentzell, there is another celebrated theory for the LDP by Donsker and Varadhan [44, 45, 46, 48]. The main difference between them is that the Freidlin-Wentzell theory is about large deviations from a deterministic trajectory by small noise but the Donsker-Varadhan theory is regarding large deviations of certain process expectations for large time with the ergodic theorem.

In the present paper, we provide a trajectory-based proof for an emergent time-inhomogeneous Gaussian process in $\mathbb{R}^n$ near a deterministic trajectory under the CLT, which follows the proof for the particular case of $\mathbb{R}^1$ in the Freidlin-Wentzell's textbook [65] (The idea of proof for $\mathbb{R}^n$ was suggested in the book but without details) and we further obtain a *Lyapunov differential equation* for the covariance of this Gaussian process. In the field of statistical physics, this Lyapunov differential equation was mentioned in the works [107, 199, 69]. However, all of those previous works were based on the small noise expansion of the associated FPEs and each approach has some limitations: In [199, 69], the dynamics was restricted to one dimension; In [107], the dynamics was for elementary processes in chemical reactions. Our approach for the Lyapunov differential equation is trajectory-based without transferring the original SDE problem to the problem of perturbations of partial differential equations (PDEs) and it is applicable for rather general multi-dimensional diffusion processes.

In contradistinction to the Freidlin-Wentzell theory and the Donsker-Varadhan theory, which are both from the standpoint of trajectories of systems, there is another approach of the LDP based on the PDEs: A logarithmic transformation to the differential generator of diffusion processes was proposed by Fleming in 1978 [63] then the PDE-based approach was applied to the LDP through solving the Hamilton-Jacobian equations (HJEs) by Evans and Ishii [57] and others. Feng and Kurtz [62] generalized this approach by refining techniques on the viscosity solutions of HJEs so that the scope of applications of it is compatible with the Freidlin-Wentzell theory and the Donsker-Varadhan theory. This rigorous mathematical PDE-based approach is corresponding to the WKB method of solving FPEs, which was introduced early by theoretical physicists [112, 83]. In the present work, our analysis of stochastic limit cycles is carried out in parallel with both the small random perturbations of SDEs and the WKB approximation of the corresponding transition probability density, which can regarded as an example of a link between the trajectory-based and the PDE-based methods. The contradistinction provides a more comprehensive portrayal of the stochastic limit cycle.

6.1.3 Time-inhomogeneous Gaussian processes from a transient state to an invariant set

By relating those two methods, one of the important results obtained in this paper is the correspondence between the local Gaussian fluctuations along a deterministic path, limit cyle or not, and the curvarture of the leading order term in the WKB method near its infimum. In the early works, the connection between the CLT and the LDP of random processes can be found in the analysis of action functional for Gaussian random processes [65] and the LDP for the empirical measures of centered stationary Gaussian processes [49, 23], in which the former follows the Friedlin-Wentzell theory and the later follows the Donsker-Varadhan theory. In the present paper, our work on the analysis of the CLT and the LDP of nonlinear systems with stochastic limit cycles, from a transient state to infinite time limit on an invariant set, are beyond those theories.

Globally, from the standpoint of probability, the existence of stationary distribution in the whole space for a stochastic stable limit cycle has been proved by Holland [89]. With the WKB method, characterizations of the stationary large deviation rate function near the cycle were studied in the previous work [51, 200, 73, 124, 126]. Locally, from the standpoint of trajectories, the dynamics is attracted to an invariant set but still capable of escaping from the set due to the multi-dimensional fluctuations except the part tangential to the cycle. In the long run, the Gaussian fluctuations along the direction tangential to the cycle is eventually smeared out and the rest of fluctuations in the hyperplane perpendicular to the cycle are outward and damped out by the dissipation toward the limit cycle [114].

In this paper, equipped with the Lyapunov equation for the covariance of the time-inhomogeneous Gaussian process, we characterize the fluctuations along the limit cycle by *asymptotic analysis*. Via a careful study of the interchange of limits of time $t \to \infty$ and $\epsilon \to 0$, with a coordinate transformation and dimension reduction on the cycle, we show that the Lyapunov equation becomes a $n-1 \times n-1$ *periodic Riccati differential equation* [19, 146, 29, 221] and the solution of equation is a positive definite matrix. We further characterize the curvature of large deviation rate function on the limit cycle by the correspondence between the covariance matrix and the curvature established in our theory.

The importance of stochastic limit cycle oscillations in physics was emphasized by Keizer [107] and van Kampen [199]. In their books, specific examples with careful studies were provided but a general analysis was missing. Our analysis by both the trajectory-based and probability-based methods, from a transient state to an invariant set, helps us to paint a clear picture of dynamics in different scopes of space and time. Additionally, the present work can be regarded as an extension of the linear approximation theory of a stochastic nonlinear system with a fixed point as the steady state, which can be found in [156, 117, 160], to an invariant set in $\mathbb{R}^n$.

Furthermore, we want to point out that stochastic limit cycles have been widely studied and applied to biological systems in two different scenarios: (1) the finite fluctuation results [208, 209, 210, 213, 60, 61] and (2) the zero noise limit with the WKB large deviation [219, 213, 218, 214]. Our work includes the following new results beyond the scopes of those two: First, we apply CLT to describe the local fluctuations around the deterministic trajectory. It provides us a new scope to investigate the limiting behavior of finite noise dynamics around its most probable path, and this scale is different from the scale of WKB large deviation. Second, for the WKB method, we include not only the large-deviation rate function but also the prefactor, which is the next order to the leading order of the large-deviation rate function. In the present work, we show that the prefactor plays an important role in stochastic limit cycles since the leading order term vanishes on the deterministic trajectory of limit cycles.

6.1.4 Organization of the paper

In Sec. 6.2, we start with a rather general small-noise diffusion process represented by a sequence of non-linear multidimentional SDEs. Based on both the trajectory-based approach and the WKB method, key lemmas regarding the time-inhomogeneous Gaussain processes and a link to the large deviation rate function are provided. In Sec. 6.3, we apply the lemmas to stochastic limit cycle oscillators. This approach is distinct to the previous works [51, 200, 73]: (i) The works [51, 200] are regarding fluctuations of limit cycles in chemical systems (The former [51] focused on analysis of a stationary FPE and the later [200] applied the WKB method directly to a master equation.); (ii) The work [73] focused on the case of one-dimensional motion on a circle. In Sec. 6.4, we introduce the scaling hypothesis of diffusion processes to construct a sequence of dynamics parameterized by ϵ. This scaling hypothesis not only serves as an useful mathematical tool of asymptotic analysis but also a scientific theory to justify the origin of ϵ.

6.2 Preliminaries

As we mentioned in Sec. 6.1, both discrete chemical kinetics and continuous mechanical motions successfully depict complex systems via introducing uncertainty. Based on the probability theory, the former is conventionally characterized by Markov jump processes (continuous-time and discrete-state) with the corresponding transition probability captured by master equations, and the later is popularly described by diffusion processes (continuous-time and continuous-state) with

SDEs. By introducing a parameter of the size of systems, at proper scales, both representations have their corresponding FPEs of the transition probability density and the HJEs of the large deviation rate function. In the present work, based on the continuous representation, we follow the direction SDE - FPE - HJE in a sequence.

6.2.1 Expansion in powers of a small parameter for diffusion processes

Let us start from a sequence of SDEs defined in Eq. (6.3)

$$\mathrm{d}\mathbf{X}_\epsilon(t) = \mathbf{b}(\mathbf{X}_\epsilon)\mathrm{d}t + [2\epsilon\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}(t), \tag{6.5}$$

and by the LLN, it converges to the following ordinary differential equation (ODE) as $\epsilon \to 0$

$$\mathrm{d}\mathbf{x}(t) = \mathbf{b}(\mathbf{x})\mathrm{d}t. \tag{6.6}$$

Let $\hat{\mathbf{x}}(t)$ be the solution of this ODE with a given initial condition $\hat{\mathbf{x}}(0) = \hat{\mathbf{x}}_0$.

We shall note that a direct application of small noise expansions for the process (6.5) by $\mathbf{X}_\epsilon(t) = \sum_{i=0}^{n} \epsilon^i \mathbf{X}_i(t)$ may fail for certain types of drift functions $\mathbf{b}$ [69]. Therefore, we need to expand $\mathbf{X}_\epsilon(t)$ with a proper scale: By the scale of the CLT, we can define a random process $\mathbf{Z}_\epsilon(t)$ near the deterministic trajectory $\hat{\mathbf{x}}(t)$

$$\mathbf{Z}_\epsilon(t) \equiv \frac{\mathbf{X}_\epsilon(t) - \hat{\mathbf{x}}(t)}{\sqrt{\epsilon}} \tag{6.7}$$

and substitute Eq. (6.7) into Eq. (6.5), we can derive that

$$\begin{aligned} & \mathrm{d}\mathbf{X}_\epsilon(t) = \mathrm{d}\hat{\mathbf{x}}(t) + \sqrt{\epsilon}\mathrm{d}\mathbf{Z}_\epsilon(t) = \mathbf{b}(\mathbf{X}_\epsilon)\mathrm{d}t + [2\epsilon\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}(t) + O(\epsilon) \\ \Rightarrow \quad & \mathrm{d}\mathbf{X}_\epsilon(t) = \mathrm{d}\hat{\mathbf{x}}(t) + \sqrt{\epsilon}\mathrm{d}\mathbf{Z}_\epsilon(t) = \left(\mathbf{b}(\hat{\mathbf{x}}) + \sqrt{\epsilon}\mathbf{A}(\hat{\mathbf{x}})\mathbf{Z}_\epsilon\right)\mathrm{d}t + [2\epsilon\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}(t) + O(\epsilon) \\ \Rightarrow \quad & \mathrm{d}\mathbf{Z}_\epsilon(t) = \mathbf{A}(\hat{\mathbf{x}})\mathbf{Z}_\epsilon\mathrm{d}t + [2\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}(t) + O(\sqrt{\epsilon}), \end{aligned} \tag{6.8}$$

where $\mathbf{A}(\hat{\mathbf{x}}(t))$ is the Jacobian matrix of $\mathbf{b}(\mathbf{x})$ evaluated at $\mathbf{x} = \hat{\mathbf{x}}(t)$. We then follow the usual approach [65] of perturbation theory to obtain an expansion in powers of the small parameter $\sqrt{\epsilon}$

$$\mathbf{Z}_\epsilon(t) = \mathbf{Z}(t) + \sqrt{\epsilon}\mathbf{Z}^{(1)}(t) + \cdots + \sqrt{\epsilon}^k\mathbf{Z}^{(k)}(t) + \cdots. \tag{6.9}$$

Apply the expansion of $\mathbf{Z}_\epsilon(t)$ in Eq. (6.9) to its SDE in Eq. (6.8), we can obtain a SDE for the zeroth approximation of $\mathbf{Z}_\epsilon(t)$

$$\mathrm{d}\mathbf{Z}(t) = \mathbf{A}(\hat{\mathbf{x}})\mathbf{Z}\mathrm{d}t + [2\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}(t). \tag{6.10}$$

The following lemma is about the solution of $\mathbf{Z}(t)$:

Lemma 6.2.1. *If each element of the Jacobian matrix $\mathbf{A}(\hat{\mathbf{x}}(t))$ is continuous for all $t \geq 0$, then for every $t > 0$, $\mathbf{Z}(t)$ is a Gaussian random variable $\mathbf{Z}(t) \sim \mathcal{N}(\boldsymbol{\mu}(t), \boldsymbol{\Sigma}(t))$ with*

$$\frac{\mathrm{d}\boldsymbol{\mu}(t)}{\mathrm{d}t} = \mathbf{A}(\hat{\mathbf{x}})\boldsymbol{\mu}, \qquad \boldsymbol{\mu}(0) = \hat{\boldsymbol{\mu}}_0, \tag{6.11}$$

$$\frac{\mathrm{d}\boldsymbol{\Sigma}(t)}{\mathrm{d}t} = \mathbf{A}(\hat{\mathbf{x}})\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{A}(\hat{\mathbf{x}})^T + 2\mathbf{D}, \qquad \boldsymbol{\Sigma}(0) = \hat{\boldsymbol{\Sigma}}_0, \tag{6.12}$$

where $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are given initial conditions.

Proof. Under the assumption that each element of $\mathbf{A}(\hat{\mathbf{x}}(t))$ is continuous for all $t \geq 0$, there exits a fundamental matrix $\mathbf{M}(t) \in \mathbb{R}^n \times \mathbb{R}^n$ satisfied the linear homogenous ordinary differential equation

$$\mathrm{d}\mathbf{M}(t) = \mathbf{A}(\hat{\mathbf{x}})\mathbf{M}\mathrm{d}t \quad \text{for all } t > 0. \tag{6.13}$$

Let $\mathbf{Z}_0$ be the given initial condition for the dynamics (6.10), we can verify the equation

$$\mathbf{Z}(t) = \mathbf{M}(t)\left(\mathbf{Z}_0 + \int_0^t \mathbf{M}^{-1}(s)[2\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}(s)\right) \quad \text{for all } t > 0, \tag{6.14}$$

by differentiating the both sides of it with the Itô lemma and Eq. (6.10) and Eq. (6.13) as follows

$$\begin{aligned}
&\mathrm{d}\left[\mathbf{M}(t)\left(\mathbf{Z}_0 + \int_0^t \mathbf{M}^{-1}(s)[2\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}(s)\right)\right] \\
&= \mathbf{M}\mathrm{d}\left(\mathbf{Z}_0 + \int_0^t \mathbf{M}^{-1}[2\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}\right) + \mathrm{d}\mathbf{M}\left(\mathbf{Z}_0 + \int_0^t \mathbf{M}^{-1}[2\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}\right) + \mathrm{d}\mathbf{M}\mathrm{d}\left(\mathbf{Z}_0 + \int_0^t \mathbf{M}^{-1}[2\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}\right) \\
&= [2\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B} + \mathbf{A}(\hat{\mathbf{x}})\mathbf{M}\mathrm{d}t\mathbf{M}^{-1}\mathbf{Z} + \mathbf{A}(\hat{\mathbf{x}})\mathbf{M}\mathrm{d}t\mathbf{M}^{-1}[2\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B} \\
&= [2\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B} + \mathbf{A}(\hat{\mathbf{x}})\mathbf{M}\mathbf{M}^{-1}\mathbf{Z}\mathrm{d}t \\
&= \mathrm{d}\mathbf{Z}(t).
\end{aligned}$$

By Eq. (6.14), for any constant vector $\mathbf{a} \in \mathbb{R}^n$, we have that

$$\mathbf{a}^T\mathbf{Z}(t) = \mathbf{a}^T\mathbf{M}(t)\left(\mathbf{Z}_0 + \int_0^t \mathbf{M}^{-1}(s)[2\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}(s)\right) = \sum_{i=1}^{n}\int_0^t f_i(s)\mathrm{d}B_i(s), \tag{6.15}$$

where $B_i \in \mathbb{R}^1$ is a collection independent and identically distributed random variables from the standard Brownian motion $\mathbf{B} = (B_1, B_2, \cdots, B_n)$, and $f_i : \mathbb{R}^1 \to \mathbb{R}^1$ is a collection of deterministic functions. Therefore, $\mathbf{a}^T\mathbf{Z}(t)$ has to be a one-dimensional Gaussian random variable since it is a linear combination of a collection of independent one-dimensional Gaussian random variables. Furthermore, by [192], using the moment generating functions, arbitrary linear combinations of the random vector $\mathbf{Z}(t)$ being an univariate Gaussian random variable implies that $\mathbf{Z}(t)$ is a multivariate Gaussian random variable.

Next, we want to find expressions of the mean and the covariance of $\mathbf{Z}(t)$ for every t. By the property of the standard Brownian motion, given a matrix $\mathbf{F}(t)$ which is independent of $\mathbf{B}$, we have that

$$\mathbb{E}\left[\int_0^t \mathbf{F}(s)\mathrm{d}\mathbf{B}(s)\right] = \mathbf{0}, \quad \text{for all } t \geq 0. \tag{6.16}$$

With (6.14) and (6.16), the first and second moment of $\mathbf{Z}(t)$ should satisfy

$$\begin{aligned}
\mathbb{E}[\mathbf{Z}(t)] &= \mathbf{M}(t)\mathbf{Z}_0 \\
\mathbb{E}[\mathbf{Z}(t)\mathbf{Z}(t)^T] &= \mathbb{E}\left[\mathbf{M}(t)\left(\mathbf{Z}_0 + \int_0^t \mathbf{M}^{-1}[2\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}\right)\left(\mathbf{Z}_0 + \int_0^t \mathbf{M}^{-1}[2\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}\right)^T \mathbf{M}(t)^T\right] \\
&= \mathbf{M}(t)\mathbf{Z}_0\mathbf{Z}_0^T\mathbf{M}(t)^T + \mathbf{M}(t)\mathbb{E}\left[\left(\int_0^t \mathbf{M}^{-1}[2\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}\right)\left(\int_0^t \mathbf{M}^{-1}[2\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}\right)^T\right]\mathbf{M}(t)^T \\
&= \mathbf{M}(t)\mathbf{Z}_0\mathbf{Z}_0^T\mathbf{M}(t)^T + 2\mathbf{M}(t)\left(\int_0^t \mathbf{M}^{-1}(s)\mathbf{D}\mathbf{M}^{-T}(s)\mathrm{d}s\right)\mathbf{M}(t)^T, \qquad (6.17)
\end{aligned}$$

where we applied the Itô isometry to the last equation. Since $\boldsymbol{\mu}(t) = \mathbb{E}[\mathbf{Z}(t)] = \mathbf{M}(t)\mathbf{Z}_0$, and $\mathbf{\Sigma}(t) = \mathbb{E}[\mathbf{Z}(t)\mathbf{Z}(t)^T] - \mathbb{E}[\mathbf{Z}(t)]\mathbb{E}[\mathbf{Z}(t)^T] = 2\mathbf{M}(t)\left(\int_0^t \mathbf{M}^{-1}(s)\mathbf{D}\mathbf{M}^{-T}(s)\mathrm{d}s\right)\mathbf{M}(t)^T$, by taking derivatives of them with respect to time, we thus obtain dynamics of $\boldsymbol{\mu}(t)$ and $\mathbf{\Sigma}(t)$ as follows

$$\frac{\mathrm{d}\boldsymbol{\mu}(t)}{\mathrm{d}t} = \mathbf{A}(\hat{\mathbf{x}})\boldsymbol{\mu}, \qquad \boldsymbol{\mu}(0) = \hat{\boldsymbol{\mu}}_0, \tag{6.18}$$

$$\frac{\mathrm{d}\mathbf{\Sigma}(t)}{\mathrm{d}t} = \mathbf{A}(\hat{\mathbf{x}})\mathbf{\Sigma} + \mathbf{\Sigma}\mathbf{A}(\hat{\mathbf{x}})^T + 2\mathbf{D}, \qquad \mathbf{\Sigma}(0) = \hat{\mathbf{\Sigma}}_0, \tag{6.19}$$

where $\hat{\boldsymbol{\mu}}_0$ and $\hat{\boldsymbol{\Sigma}}_0$ are given initial conditions.

□

By Lemma 6.2.1, we obtained a time-inhomogeneous Gaussian process from a multi-dimensional nonlinear diffusion process at the scale of the CLT with the covariance captured by the Lyapunov differential equation (6.19). Additionally, this lemma can be applied to solve the FPE

$$\frac{\partial p_\epsilon}{\partial t} = -\nabla \cdot \mathbf{J}[p_\epsilon], \quad \mathbf{J}[p_\epsilon] \equiv \mathbf{b}(\mathbf{x})p_\epsilon - \epsilon \mathbf{D}\nabla p_\epsilon \tag{6.20}$$

with certain boundary conditions. Since the function $\mathbf{b}$ is nonlinear and multi-dimensional, this type of PDE problems may not be easy to solve directly. Under certain conditions [65, 101], the diffusion process (6.5) is associated with this FPE. By expanding the solution of the FPE (6.20) and the diffusion process (6.5) respectively

$$p_\epsilon(\mathbf{x}, t) = \frac{1}{\sqrt{\epsilon}}\hat{p}_\epsilon(\mathbf{z}, t) \quad \text{and} \quad \hat{p}_\epsilon(\mathbf{z}, t) = \hat{p}_0(\mathbf{z}, t) + \sum_{n=1}^{\infty}(\sqrt{\epsilon})^n \hat{p}_n(\mathbf{z}, t), \tag{6.21}$$

$$\mathbf{Z}_\epsilon(t) = \frac{\mathbf{X}_\epsilon(t) - \hat{\mathbf{x}}(t)}{\sqrt{\epsilon}} \quad \text{and} \quad \mathbf{Z}_\epsilon(t) = \mathbf{Z}(t) + \sum_{n=1}^{\infty}(\sqrt{\epsilon})^n \mathbf{Z}^{(n)}(t), \tag{6.22}$$

we can check that $\hat{p}_0(\mathbf{z}, t)$ is the probability density of $\mathbf{Z}(t)$. Following from Lemma 6.2.1, we thus obtain an approximation solution of the FPE by having the dynamics of the mean and covariance of $\mathbf{Z}(t)$. To transform the FPE problem into the problem of a diffusion process, the above example is an application of Lemma 6.2.1 to attack a complicated boundary value problem.

6.2.2 Approximations by the asymptotic theory akin to the WKB

In Sec. 6.1.1, we pointed out that using the Kramers-Moyal Fokker-Planck equation for a mater equation may fail in some cases. Instead, the deterministic behavior and the local fluctuation of a jump Markov process can be obtained by the Ω expansion with respect to two different scales. In addition to the Ω expansion, another approach by the WKB approximation has been applied to give a full analysis of master equations [112, 91, 200, 142]. In this method, the WKB ansatz is assumed for the solution of the Kramers-Moyal expansion of a master equation without truncating higher-order terms, so this method has no problem unlike the Kramers-Moyal Fokker-Planck equation.

As the success of the WKB approximation of master equations for Markov jump processes, this method has also been used to the associated Fokker-Planck equations of diffusion processes [83, 171]. Here we want to link our trajectory-based approach in Sec. 6.2.1 to the probability-based approach by the WKB approximation.

Recall that the path behaviors of the diffusion process is described by the n-dimensional SDE (6.5)

$$\mathrm{d}\mathbf{X}_\epsilon(t) = \mathbf{b}(\mathbf{X}_\epsilon)\mathrm{d}t + \sqrt{2\epsilon\mathbf{D}}\mathrm{d}\mathbf{B}_t. \tag{6.23}$$

In order to link the SDE (6.23) to the WKB ansatz, we need to find its probability-density representation by a FPE. From Eq. (6.20) to Eq. (6.22), we have illustrated a way to convert a PDE problem to a SDE problem; on the other hand, by the semigroup approaches [58, 101], we can also convert a SDE problem to a PDE problem. Under certain conditions [101], the original SDE problem can be characterized by the solution of the FPE

$$\frac{\partial p_\epsilon}{\partial t} = -\nabla \cdot \mathbf{J}[p_\epsilon], \quad \mathbf{J}[p_\epsilon] \equiv \mathbf{b}(\mathbf{x})p_\epsilon - \epsilon\mathbf{D}\nabla p_\epsilon. \tag{6.24}$$

We shall note that, as $\epsilon \to 0$, the FPE reduces to a first-order differential equation so the perturbation of the solution p_ϵ follows the singular perturbation theory. To attack this singular perturbation problem, we adopt an asymptotic series of p_ϵ with a proper scaled variable $\mathbf{z} = (\mathbf{x} - \hat{\mathbf{x}})/\sqrt{\epsilon}$,

$$p_\epsilon(\mathbf{x}, t) = \frac{1}{\sqrt{\epsilon}}\hat{p}_\epsilon(\mathbf{z}, t) \quad \text{and} \quad \hat{p}_\epsilon(\mathbf{z}, t) = \sum_{n=0}^{\infty}(\sqrt{\epsilon})^n \hat{p}_n(\mathbf{z}, t). \tag{6.25}$$

In parallel, there is another complete asymptotic theory for the solution of FPE akin to the WKB ansatz [112, 83]

$$p_\epsilon(\mathbf{x}, t) = a(\epsilon, t)\exp\left[-\frac{1}{\epsilon}\sum_{n=0}^{\infty}\phi_n(\mathbf{x}, t)\epsilon^n\right], \tag{6.26}$$

where $a(\epsilon, t)$ is a normalization factor. The expansion (6.26) with the series $\epsilon^{-1}\phi_0 + \phi_1 + \epsilon\phi_2 + \cdots$ was justified by the *extensive property* of $p_\epsilon(\mathbf{x}, t)$, i.e. it keeps the form (6.26) as time evolves [112].

We will connect those two types of expansions in Lemma 6.2.4. To prove the lemma, we first give two useful lemmas on the asymptotic evaluation of various integrals [13]. All the proofs can be found in Appendix 6.6.1.

Lemma 6.2.2. *For sufficiently smooth scalar functions $f(\mathbf{x})$ and $h(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$,*

$$\int_{\mathbb{R}^n} f(\mathbf{x}) e^{-h(\mathbf{x})/\epsilon} \mathrm{d}\mathbf{x} = \sqrt{\frac{2\pi\epsilon}{\det[\nabla\nabla h(\mathbf{x}^*)]}} e^{-\frac{h(\mathbf{x}^*)}{\epsilon}} \Big[f(\mathbf{x}^*) + \epsilon\eta(\mathbf{x}^*) + O\big(\epsilon^2\big)\Big], \tag{6.27}$$

$$\begin{aligned} &\frac{\displaystyle\int_{\mathbb{R}^n} f(\mathbf{x}) e^{-h(\mathbf{x})/\epsilon} \mathrm{d}\mathbf{x}}{\displaystyle\int_{\mathbb{R}^n} e^{-h(\mathbf{x})/\epsilon} \mathrm{d}\mathbf{x}} \qquad (6.28)\\ = \;& f(\mathbf{x}^*) + \epsilon \left[\frac{f''_{ij}(\mathbf{x}^*)\Xi_{ij}}{2} - \frac{f'_i(\mathbf{x}^*) h'''_{jk\ell}(\mathbf{x}^*) \Xi^{\frac{1}{2}}_{i\mu} \Xi^{\frac{1}{2}}_{j\nu} \Xi^{\frac{1}{2}}_{k\rho} \Xi^{\frac{1}{2}}_{\ell\kappa} \Theta_{\mu\nu\rho\kappa}}{6}\right] + O(\epsilon^2), \end{aligned}$$

as $\epsilon \to 0$, in which Einstein's summation rule is adopted, $\mathbf{x}^*$ is the global minimum of $h(\mathbf{x})$, and

$$\begin{aligned} \eta(\mathbf{x}^*) &= \frac{f''_{ij}(\mathbf{x}^*)\Xi_{ij}}{2} - \left[\frac{f'_i(\mathbf{x}^*) h'''_{jk\ell}(\mathbf{x}^*)}{6} + \frac{f(\mathbf{x}^*) h''''_{ijk\ell}(\mathbf{x}^*)}{24}\right] \Xi^{\frac{1}{2}}_{i\mu} \Xi^{\frac{1}{2}}_{j\nu} \Xi^{\frac{1}{2}}_{k\rho} \Xi^{\frac{1}{2}}_{\ell\kappa} \Theta_{\mu\nu\rho\kappa} \\ &+ \frac{f(\mathbf{x}^*)[h'''_{ijk}(\mathbf{x}^*)]^2}{72} \Xi^{-\frac{1}{2}}_{i\mu} \Xi^{-\frac{1}{2}}_{i\mu'} \Xi^{-\frac{1}{2}}_{j\nu} \Xi^{-\frac{1}{2}}_{j\nu'} \Xi^{-\frac{1}{2}}_{k\rho} \Xi^{-\frac{1}{2}}_{k\rho'} \Lambda_{\mu\mu'\nu\nu'\rho\rho'}. \end{aligned} \tag{6.29}$$

The covariance matrix $\boldsymbol{\Xi} = \big[\nabla\nabla h(\mathbf{x}^*)\big]^{-1}$, and the multi-indexed $\Theta_{ijk\ell}$ and $\Lambda_{\mu\mu'\nu\nu'\rho\rho'}$ are

$$\Theta_{\mu\nu\rho\kappa} = \int_{\mathbb{R}^n} \frac{y_\mu y_\nu y_\rho y_\kappa}{\big(2\pi\big)^{n/2}} \exp\left[-\frac{\mathbf{y}^T\mathbf{y}}{2}\right] \mathrm{d}\mathbf{y}, \tag{6.30}$$

$$\Lambda_{\mu\mu'\nu\nu'\rho\rho'} = \int_{\mathbb{R}^n} \frac{y_\mu y_{\mu'} y_\nu y_{\nu'} y_\rho y_{\rho'}}{\big(2\pi\big)^{n/2}} \exp\left[-\frac{\mathbf{y}^T\mathbf{y}}{2}\right] \mathrm{d}\mathbf{y}. \tag{6.31}$$

By Lemma 6.2.2, we can obtain the following lemma, which is very useful in the integrals with respect to a probability density approximated by the WKB method.

Lemma 6.2.3. *For sufficiently smooth functions $f(\mathbf{x})$, $g(\mathbf{x})$, and $h(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$,*

$$\begin{aligned} & \frac{\int_{\mathbb{R}^n} f(\mathbf{x}) g(\mathbf{x}) e^{-\frac{h(\mathbf{x})}{\epsilon}} \mathrm{d}\mathbf{x}}{\int_{\mathbb{R}^n} g(\mathbf{x}) e^{-\frac{h(x)}{\epsilon}} \mathrm{d}\mathbf{x}} \\ = \quad & f(\mathbf{x}^*) + \epsilon \left[f_i'(\mathbf{x}^*)(\log g)_i'(\mathbf{x}^*)\Xi_{ij} + \frac{f_{ij}''(\mathbf{x}^*)\Xi_{ij}}{2} - \frac{f_i'(\mathbf{x}^*) h_{jk\ell}'''(\mathbf{x}^*) \Xi_{i\mu}^{\frac{1}{2}} \Xi_{j\nu}^{\frac{1}{2}} \Xi_{k\rho}^{\frac{1}{2}} \Xi_{\ell\kappa}^{\frac{1}{2}} \Theta_{\mu\nu\rho\kappa}}{6} \right] + O(\epsilon^2), \end{aligned} \tag{6.32}$$

as $\epsilon \to 0$, in which Einstein's summation rule is adopted, $\mathbf{x}^*$ is the global minimum of $h(\mathbf{x})$,

Now, we are ready to relate the two kinds of asymptotic series. Recall that the two expansions are

$$\hat{p}_\epsilon(\mathbf{z}, t) = \sum_{n=0}^{\infty} (\sqrt{\epsilon})^n \hat{p}_n(\mathbf{z}, t), \tag{6.33}$$

$$p_\epsilon(\mathbf{x}, t) = a(\epsilon, t) \exp\left[-\frac{1}{\epsilon} \sum_{n=0}^{\infty} \phi_n(\mathbf{x}, t) \epsilon^n \right], \tag{6.34}$$

where $\mathbf{z} = (\mathbf{x} - \hat{\mathbf{x}})$ and $p_\epsilon(\mathbf{x}, t) = \hat{p}_\epsilon(\mathbf{z}, t)/\sqrt{\epsilon}$. Since we will focus on the first two orders in the WKB approximation, we specifically denote that $\phi_0 := \varphi$ and $\phi_1 := \ln \omega$. These two functions have particular meanings: φ is known as the large deviation rate function [65, 38], and ω is known as the prefactor [83, 73], or the phase space factor [199], or degeneracy in the classical statistical mechanical terminology [171].

Recall that the time-dependent matrix $\boldsymbol{\Sigma}(t)$ in Lemma 6.2.1 is the covariance matrix of $\mathbf{Z}(t)$ and we have checked that $\mathbf{Z}(t)$ has the density $\hat{p}_0(\mathbf{z}, t)$. Therefore, for $\boldsymbol{\Sigma}(t)$, it has the formula $\Sigma_{ij}(t) = \int_{\mathbb{R}^n} z_i z_j \hat{p}_0(\mathbf{z}, t) \mathrm{d}\mathbf{z}$, for $1 \le i \le n,\ 1 \le i \le n$. Here we further define a time dependent first moment vector $\mathbf{m}(t)$ with respect to $\hat{p}_1(\mathbf{x}, t)$ by $m_i(t) = \int_{\mathbb{R}^n} z_i \hat{p}_1(\mathbf{z}, t) \mathrm{d}\mathbf{z}$, for $1 \le i \le n$. Note that the functions $\hat{p}_0$ and $\hat{p}_1$ are given in the expansion (6.33). Under this framework, $\boldsymbol{\Sigma}(t)$ and $\mathbf{m}(t)$ must satisfy the differential equations

$$\frac{\mathrm{d}\boldsymbol{\Sigma}(t)}{\mathrm{d}t} = \mathbf{A}(\hat{\mathbf{x}}(t))\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{A}(\hat{\mathbf{x}}(t))^T + 2\mathbf{D}, \qquad \boldsymbol{\Sigma}(0) = \hat{\boldsymbol{\Sigma}}_0, \tag{6.35}$$

$$\frac{\mathrm{d}\mathbf{m}(t)}{\mathrm{d}t} = \mathbf{A}(\hat{\mathbf{x}}(t)) + \mathbf{H}(\hat{\mathbf{x}}(t))\boldsymbol{\Sigma}, \quad \mathbf{m}(0) = \hat{\mathbf{m}}_0, \tag{6.36}$$

in which the initial conditions $\hat{\mathbf{\Sigma}}_0$ and $\hat{\mathbf{m}}_0$ are given by the distribution of $\mathbf{X}_\epsilon(0)$. For example, if the initial probability density of $\mathbf{X}_\epsilon(0)$ is purely Gaussian, then $\hat{p}_1(\mathbf{z}, 0) = 0$ for all $\mathbf{z}$ hence $\hat{\mathbf{m}}_0 = 0$. Furthermore, $\mathbf{A}(\mathbf{x})$ is the Jacobian matrix of $\mathbf{b}(\mathbf{x})$, and $\mathbf{H}(\mathbf{x})$ is a rank 3 tensor with $\mathbf{H}_i(\mathbf{x})$ being the Hessian matrix of $b_i(\mathbf{x})$, $1 \le i \le n$. Eq. (6.35) of $\mathbf{\Sigma}(t)$ is from Lemma 6.2.1, and Eq. (6.36) can be verified by plugging the expansion (6.33) into the the Fokker-Planck equation (6.24) and using integration by parts.

Based on the above setup, $\mathbf{\Sigma}(t)$, $\mathbf{m}(t)$ are related to the expansion (6.33) and $\varphi(\mathbf{x}, t)$, $\omega(\mathbf{x}, t)$ are related to the expansion (6.34), then we have the following lemma for their correspondence. Recall that $\hat{\mathbf{x}}(t)$ is the emergent deterministic trajectory of the diffusion process $\mathbf{X}_\epsilon(t)$ as $\epsilon \to 0$.

Lemma 6.2.4. *$\mathbf{\Sigma}(t)$, $\mathbf{m}(t)$, $\varphi(\hat{\mathbf{x}}(t), t)$, and $\omega(\hat{\mathbf{x}}(t), t)$ must satisfy the equations*

$$\mathbf{\Sigma}(t) = \left[\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t)\right]^{-1}, \tag{6.37}$$

$$\mathbf{m}(t) = \left[\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t)\right]^{-1} \nabla \log \omega(\hat{\mathbf{x}}(t), t) - \frac{1}{6}\nabla\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t)\Theta, \tag{6.38}$$

where $\left(\nabla\nabla\nabla\varphi(\mathbf{x}, t)\Theta\right)_i = \sum_{jkl\mu\nu\kappa\rho} \varphi'''_{jk\ell}(\mathbf{x}, t)\Xi^{\frac{1}{2}}_{i\mu}\Xi^{\frac{1}{2}}_{j\nu}\Xi^{\frac{1}{2}}_{k\rho}\Xi^{\frac{1}{2}}_{\ell\kappa}\Theta_{\mu\nu\rho\kappa}$, $\mathbf{\Xi} = \left[\nabla\nabla\varphi(\mathbf{x}, t)\right]^{-1}$*, and* Θ *is defined by Eq.* (6.30) *in Lemma 6.2.2.*

Proof. By the change of variable $\mathbf{z} = (\mathbf{x} - \hat{\mathbf{x}})/\sqrt{\epsilon}$, we have the following two equations

$$\int (\sqrt{\epsilon}\mathbf{z})(\sqrt{\epsilon}\mathbf{z})^T \hat{p}_\epsilon(\mathbf{z}, t)\mathrm{d}\mathbf{z} = \int (\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T p_\epsilon(\mathbf{x}, t)\mathrm{d}\mathbf{x}, \tag{6.39}$$

$$\int (\sqrt{\epsilon}\mathbf{z})\hat{p}_\epsilon(\mathbf{z}, t)\mathrm{d}\mathbf{z} = \int (\mathbf{x} - \hat{\mathbf{x}})p_\epsilon(\mathbf{x}, t)\mathrm{d}\mathbf{x}. \tag{6.40}$$

Plug the expansion (6.33) into the left side of (6.39), by Lemma 6.2.1, the left side of (6.39) becomes

$$\epsilon\mathbf{\Sigma}(t) + o(\epsilon). \tag{6.41}$$

For a fixed t, the point $\mathbf{x} = \hat{\mathbf{x}}(t)$ on the deterministic trajectory is the global minimum of $\varphi(\mathbf{x}, t)$. Therefore, to plug the expansion (6.34) into the right side of (6.39), by Lemma 6.2.3, we have

$$\epsilon\left[\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t)\right]^{-1} + o(\epsilon). \tag{6.42}$$

With Eq. (6.41) and Eq. (6.42), we thus obtain

$$\mathbf{\Sigma}(t) = [\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t)]^{-1}. \tag{6.43}$$

By a similar approach, applying Lemma 6.2.1 to the left side of (6.40) and Lemma 6.2.3 to the right side of it, we have that

$$\mathbf{m}(t) = [\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t)]^{-1} \nabla \ln \omega(\hat{\mathbf{x}}(t), t) - \frac{1}{6}\nabla\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t)\Theta. \tag{6.44}$$

□

The leading order $\varphi(\mathbf{x}, t)$ of the time-dependent WBK ansatz (6.26) is known as a time-dependent large deviation rate function. Our work (Lemma 6.2.4) relates the curvature of the time-dependent large deviation rate function near its infimum with the local Gaussian fluctuations of diffusion processes. In the case of independent and identically distributed (i.i.d.) random variables sampling, the inverse of the curvature of large deviation rate function equivalent to the variance of each random variable is one of the important properties of the rate function [22, 193]. Eq. (6.37) in Lemma 6.2.4 can be regarded as an extension of this property to the case of random processes. The Freidlin-Wentzell theory [65] gives a clear definition of the large deviation rate function of random processes. From the trajectory standpoint, the action functional is defined as [65]

$$\mathcal{S}_{0,t}(\xi) = \frac{1}{4}\int_0^t [\dot{\xi}_s - \mathbf{b}(\xi_s)]\mathbf{D}^{-1}[\dot{\xi}_s - \mathbf{b}(\xi_s)]\mathrm{d}s, \tag{6.45}$$

where ξ is the set of all smooth paths of the process (6.5) on the interval $[0, t]$. Then the time-dependent large deviation rate function is the minimum of action among the set of ξ

$$\varphi(\mathbf{x}, t) = \min_{\xi_0=\mathbf{x}_0, \xi_t=\mathbf{x}} \mathcal{S}_{0,t}(\xi), \tag{6.46}$$

in which $\mathbf{x}_0$ and $\mathbf{x}$ are the initial and end conditions of the process, respectively.

By Eqs. (6.45) and (6.46), $\varphi(\mathbf{x}, t)$ is no longer just a mathematical concept of the leading order term of the logarithmic asymptotics of probability densities. Borrowing the concept from classical mechanics, the integrand in the integral (6.45) is called the Lagrangian of the action and there is a corresponding Hamiltonian of the system defined by the Legendre dual of the Lagrangian [222]

$$H(\xi, \mathbf{p}) = \mathbf{b}(\xi) \cdot \mathbf{p} + \mathbf{D}\mathbf{p} \cdot \mathbf{p}. \tag{6.47}$$

Furthermore, based on the Hamiltonian given in Eq. (6.47), the large deviation rate function has to satisfy the Hamilton-Jacobi equation

$$\frac{\partial \varphi(\mathbf{x}, t)}{\partial t} = -H(\mathbf{x}, \nabla\varphi). \tag{6.48}$$

Finding solutions of the Hamilton-Jacobi equation (6.48) is still an open problem. By Lemma 6.2.4, with the dynamics of $\mathbf{\Sigma}(t)$ in Eq. (6.35) and $\mathbf{m}(t)$ in Eq. (6.36), if the solution of the prefactor $\omega(\mathbf{x}, t)$ is given, e.g., an uniform prefactor, then we can derive the dynamics of $\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t)$ and $\nabla\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t)$. These results can help us approximate the solution of the HJE near its infimum: by the multivariable Taylor's expansion, for $\|\mathbf{x} - \hat{\mathbf{x}}\| < \delta$, we have a *third order approximation*

$$\varphi(\mathbf{x}, t) = \frac{1}{2}\left[(\mathbf{x} - \hat{\mathbf{x}}(t)) \cdot \nabla\right]^2 \varphi(\hat{\mathbf{x}}(t), t) + \frac{1}{3}\left[(\mathbf{x} - \hat{\mathbf{x}}(t)) \cdot \nabla\right]^3 \varphi(\hat{\mathbf{x}}(t), t) + o(\delta^3), \tag{6.49}$$

and the first two terms on the right side can be numerically solved with the dynamics of $\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t)$ and $\nabla\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t)$ obtained by our theory.

To summarize the novelty and the significance of Lemma 6.2.4 in the following three points:

1. We show rigorously that the covariance matrix of the local time-inhomogeneous Gaussian process near a deterministic trajectory is equivalent to the inverse of the curvature of the time-dependent large deviation rate function near its infimum.

2. By having the dynamics of $\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t)$ and $\nabla\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t)$ from the lemma, we can obtain a third-order approximation of the solution of the HJE (6.48) near its infimum.

3. For analyzing a stochastic stable limit cycle in the next section, with the Lyapunov differential equation of $\mathbf{\Sigma}(t)$ (6.35) and the equation $\mathbf{\Sigma}(t) = \left[\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t)\right]^{-1}$ in the lemma, we can further study the asymptotic behaviors of the time-inhomogeneious Gaussian process from a transient state to an invariant set, and relate this result to the previous works

[200, 73, 124, 126] on the curvature of the stationary large deviation rate function near a limit cycle.

6.3 Main results of stochastic limit-cycle oscillations

In this section, we focus on nonlinear stochastic complex systems having stable limit cycles at the macroscopic scale. In chapter XIII. 7 of the textbook by van Kampen [199], the author proposed two examples of stochastic systems with stable limit cycles: one is dynamics of the Brusselator in the chemical reaction and the other one is the generalized Ginzburg–Landau equation in statistical mechanics. The model of the former started from a master equation for the Markov jump process and the later began with a SDE for the diffusion process. More examples of stochastic chemical kinetics with limit cycle oscillators were thoroughly discussed in the previous works [51, 200]. In contradistinction to stochastic chemical kinetics, our work follows the idea of the second example in the van Kampen's book along the line of the continuous representation of complex systems: we start with a randomly perturbed diffusion process satisfied the sequence of SDEs (6.5). Recall it has the form

$$\mathrm{d}\mathbf{X}_\epsilon(t) = \mathbf{b}(\mathbf{X}_\epsilon)\mathrm{d}t + [2\epsilon\mathbf{D}]^{\frac{1}{2}}\mathrm{d}\mathbf{B}(t). \tag{6.50}$$

In this section, we assume $\mathbf{D}$ is positive definite in particular. Furthermore, there is an emergent deterministic dynamics as $\epsilon \to 0$,

$$\mathrm{d}\mathbf{x}(t) = \mathbf{b}(\mathbf{x})\mathrm{d}t, \tag{6.51}$$

and the solution $\hat{\mathbf{x}}(t)$ of this ODE (6.51) with initial condition $\hat{\mathbf{x}}(0)$ has a invariant solution as a stable limit cycle Γ. Recall that the corresponding FPE of Eq. (6.50) is

$$\frac{\partial p_\epsilon}{\partial t} = -\nabla \cdot \mathbf{J}[p_\epsilon], \quad \mathbf{J}[p_\epsilon] \equiv \mathbf{b}(\mathbf{x})p_\epsilon - \epsilon\mathbf{D}\nabla p_\epsilon. \tag{6.52}$$

To analyze stochastic limit cycles defined by Eq. (6.50) - Eq. (6.52) requires asymptotic analysis involving two limits ($\epsilon \to 0$ and $t \to \infty$). Therefore, in the first part of Sec. 6.3.1, we provide a brief review of the previous work on asymptotic analysis of Eq. (6.53) and Eq. (6.54). We shall note that Eq. (6.53) is a HJE for $\varphi(\mathbf{x}, t)$ and Eq. (6.54) is a continuity equation for $\omega(\mathbf{x}, t)$. This pair of equations was known in the previous work by applying WKB method to the semi-classical limit of Schrödinger equations [85, 12]. In the second part of Sec. 6.3.1, we show a new result for stochastic limit cycles (Theorem 6.3.1) by asymptotic analysis of Lemmas 6.2.1 and 6.2.4 from a finite time to the infinite time limit.

6.3.1 The process asymptotic to a time-invariant set

In Sec. 6.2.2, we relate the time-dependent large deviation rate function and the prefactor in the WKB method with the stochastic trajectories of randomly perturbed dynamics systems. By this correspondence, we further inspect asymptotic behaviors of the time-dependent large deviation rate function and the prefactor as time goes to infinity. By plugging the WKB ansatz (6.26) into the FPE (6.24) with equating likeorder terms, we obtain two partial differential equations of $\varphi(\mathbf{x}.t)$ and $\omega(\mathbf{x}, t)$ respectively,

$$\frac{\partial \varphi(\mathbf{x}, t)}{\partial t} = -\nabla\varphi(\mathbf{x}, t) \cdot \boldsymbol{\gamma}(\mathbf{x}, t) \tag{6.53}$$

$$\frac{\partial \omega(\mathbf{x}, t)}{\partial t} = -\nabla \cdot (\boldsymbol{\gamma}(\mathbf{x}, t)\omega(\mathbf{x}, t)) - \mathbf{D}\nabla\varphi(\mathbf{x}, t) \cdot \nabla\omega(\mathbf{x}, t), \tag{6.54}$$

where $\boldsymbol{\gamma}(\mathbf{x}, t) = \mathbf{D}\nabla\varphi(\mathbf{x}, t) + \mathbf{b}(\mathbf{x})$. We will later discuss the meaning of $\boldsymbol{\gamma}(\mathbf{x}, t)$, which is closely related the concept of probability flux and the Onsager's thermodynamic force.

Let us start from analysis of the solution of $\varphi(\mathbf{x}, t)$ in Eq. (6.53), which is a Hamilton-Jacobi equation. The HJE derived in this place by equating the leading order term in the time-dependent WKB ansatz is the same one obtained in the previous derivation (Eq. (6.46) - Eq. (6.48)) by minimizing of action among all possible smooth paths. However, when we further analyze time-invariant solutions of this HJE, we shall notice an essential difference between the WKB-type method and the trajectory-based method, which is due to the interchange of limits of t and ϵ.

Following the idea of WKB anstaz (6.26) for the time-dependent probability density, if the invariant probability exists, it could be written as the asymptotic form

$$\pi_\epsilon(\mathbf{x}) = \hat{a}(\epsilon) \exp\left[-\frac{1}{\epsilon}\hat{\varphi}(\mathbf{x}) + \ln\omega(\mathbf{x}) + O(\epsilon)\right] \tag{6.55}$$

and it is equivalent to say that

$$\hat{\varphi}(\mathbf{x}) = -\lim_{\epsilon\to 0}\lim_{t\to\infty} \ln p_\epsilon(\mathbf{x}, t|\mathbf{x}_0, 0), \tag{6.56}$$

where $\hat{\varphi}$ is independent of the initial condition $\mathbf{x}_0$ and is well-defined in the whole space $\mathbb{R}^n$. On the other hand, from the standpoints of trajectories, the quasipotential of the system is defined by

$$\varphi(\mathbf{x}; \mathbf{x}_f) := \inf_{t>0} \inf_{\xi} \{\mathcal{S}_{0,t}(\xi) : \xi_0 = \mathbf{x}_f, \xi_t = \mathbf{x}\}, \tag{6.57}$$

where $\mathbf{x}_f$ is a fixed point of the deterministic dynamics $\hat{\mathbf{x}}' = \mathbf{b}(\hat{\mathbf{x}})$. This definition extends the definition of minimizing the action (6.46) from a fixed t to all $t > 0$ and it has a corresponding probabilistic representation [222]

$$\varphi(\mathbf{x}; \mathbf{x}_f) = -\lim_{t\to\infty}\lim_{\epsilon\to 0} \ln p_\epsilon(\mathbf{x}, t|\mathbf{x}_f, 0). \tag{6.58}$$

The quasipotential (6.57) is one of the invariant solutions of the HJE (6.53) [65]. It may contain non-differential parts since the HJE can have a non-smooth solution after a certain finite time by studying the characteristics of it [56]. We shall emphasize that the two potentials, $\hat{\varphi}(\mathbf{x})$ and $\varphi(\mathbf{x}; \mathbf{x}_f)$, have the same shape only in the domain (denoted by $\mathcal{D}$) where $\hat{\varphi}(\mathbf{x}) > \hat{\varphi}(\mathbf{x}_f)$ and $\hat{\varphi}(\mathbf{x})$ is continuously differentiable with $\nabla\hat{\varphi}(\mathbf{x}) \neq 0$ by the Freidlin-Wentzell uniqueness theorem of the orthogonal decomposition of the drift function $\mathbf{b}$ [65]. Since the quasipotential $\varphi(\mathbf{x}; \mathbf{x}_f)$ is defined strictly by the trajectories of dynamics in Eq. (6.57), to equate $\hat{\varphi}(\mathbf{x})$ and $\varphi(\mathbf{x}; \mathbf{x}_f)$ on $\mathcal{D}$, we can justify the leading order term of the WKB "ansatz" (6.55) for the invariant probability at least on $\mathcal{D}$. Analogously, $\mathbf{x}_f$ can be extended from a fixed point to a invariant set, so the above statement is also true for the systems with a stable limit cycle [65]. In the following work, we will restrict our analysis of dynamics contained in $\mathcal{D}$ and assume $\mathcal{D}$ is compact, and use one brief notation $\varphi(\mathbf{x})$ to represent both of the potentials.

Based on the above justification of the time-invariant WKB ansatz (6.26), we can plug it into the stationary FPE (6.52) to get the system satisfied three equations [171]

$$\mathbf{b}(\mathbf{x}) = -\mathbf{D}\nabla\varphi(\mathbf{x}) + \boldsymbol{\gamma}(\mathbf{x}), \tag{6.59}$$

$$\nabla\varphi(\mathbf{x}) \cdot \boldsymbol{\gamma}(\mathbf{x}) = 0, \tag{6.60}$$

$$\nabla \cdot (\omega(\mathbf{x})\boldsymbol{\gamma}(\mathbf{x})) = -\nabla\varphi(\mathbf{x}) \cdot \mathbf{D}\nabla\omega(\mathbf{x}). \tag{6.61}$$

Note that the vector field $\mathbf{b}(\mathbf{x})$ represents deterministic dynamics and can be decomposed to two terms $\nabla\varphi(\mathbf{x}) \perp \boldsymbol{\gamma}(\mathbf{x})$ which is consistent with the FW's orthogonal decomposition [65]. In comparison to the gradient flow $\nabla\varphi(\mathbf{x})$, dynamics following the vector field $\boldsymbol{\gamma}$ represents the part of circular motion of $\mathbf{b}$. By the previous works [89, 114, 200, 73, 124, 126], we have the following three propositions of $\varphi(\mathbf{x})$:

1. We have justified the leading order term $\varphi(\mathbf{x})$ in the stationary WKB ansatz. But it is based on the existence of the stationary probability distribution π_ϵ. The existence of π_ϵ for stochastic stable limit cycles has been proved in the work of Holland [89].

2. On a limit cycle, $\varphi(\mathbf{x})$ and $\nabla\varphi(\mathbf{x})$ are always zero. The landscape of $\varphi(\mathbf{x})$ has a Mexican hat shape and the bottom of the Mexican hat ring characterizes the deterministic trajectory of the cycle [124]; And the second derivative of $\varphi(\mathbf{x})$ tangential to the cycle is also zero, which means the Gaussian fluctuations along the direction tangential to the cycle is eventually smeared out [114, 200, 73].

3. Along the cycle, the matrix $\nabla\nabla\varphi(\mathbf{x})$ is positive semi-definite [200, 126]. Specifically, the smallest eigenvalue of $\nabla\nabla\varphi(\mathbf{x})$ is zero on the cycle and the corresponding eigenvector is tangential to the cycle (by the proposition (2)). The rest of eigenvalues are positive, i.e., the Gaussian fluctuations perpendicular to the cycle are outward and damped out by the dissipation toward the limit cycle.

The above is the setup of $\varphi(\mathbf{x}, t)$. Let us continue on analysis of the solution of $\omega(\mathbf{x}, t)$ in Eq. (6.54). It can be rearranged as

$$\frac{\partial\omega(\mathbf{x}, t)}{\partial t} + \nabla\cdot(\mathbf{b}(\mathbf{x})\omega(\mathbf{x}, t)) = -2\mathbf{D}\nabla\varphi(\mathbf{x}, t)\nabla\omega(\mathbf{x}, t) - \mathbf{D}\nabla\nabla\varphi(\mathbf{x}, t)\omega(\mathbf{x}, t), \tag{6.62}$$

where $\mathbf{D}\nabla\nabla\varphi(\mathbf{x}, t)$ is the the Frobenius product of the matrix $\mathbf{D}$ and the matrix $\nabla\nabla\varphi(\mathbf{x}, t)$. We can identify Eq. (6.62) is a *continuity equation* that describes the transport following the vector field $\hat{\mathbf{x}}' = \mathbf{b}(\hat{\mathbf{x}})$ with a density-dependent sink (source) term on the right hand side. Therefore, the solution of Eq. (6.62) gives us a measure of a large number particle system “without” noise by the

Eulerian description of dynamics $\hat{\mathbf{x}}' = \mathbf{b}(\hat{\mathbf{x}})$. The original effect of noise, $\mathbf{D}$, appears in the sink (source) term in this continuity equation of ω. On the other hand, if we follow the dynamics of $\hat{\mathbf{x}}' = \mathbf{b}(\hat{\mathbf{x}})$, we have the dynamics of ω by the Lagrangian description

$$\begin{aligned}\frac{\mathrm{d}\omega(\hat{\mathbf{x}}(t),t)}{\mathrm{d}t} &= \frac{\partial\omega(\hat{\mathbf{x}}(t),t)}{\partial t} + \nabla\omega(\hat{\mathbf{x}}(t),t)\frac{\mathrm{d}\hat{\mathbf{x}}(t)}{\mathrm{d}t} \\ &= \nabla\cdot\mathbf{b}(\mathbf{x})\omega(\hat{\mathbf{x}}(t),t) - \mathbf{D}\nabla\nabla\varphi(\hat{\mathbf{x}}(t),t)\omega(\hat{\mathbf{x}}(t),t) \\ &= -\nabla\cdot\boldsymbol{\gamma}(\mathbf{x}(t),t)\omega(\hat{\mathbf{x}}(t),t).\end{aligned} \tag{6.63}$$

To compare Eq. (6.62) with Eq. (6.61), we have that $\omega(\mathbf{x})$ in the WKB ansatz (6.55) is one of invariant solutions of Eq. (6.62). However, in distinction to the HJE of $\varphi(\mathbf{x},t)$, the uniqueness and the smoothness of invariant solutions of the PDE (6.62) are not discussed in the present work.

Based on the above setup, we have the following theorem about the curvature of invariant large deviation rate function $\nabla\nabla\varphi(\mathbf{x})$ and the logarithm of the prefactor $\ln\omega(\mathbf{x})$ along the dynamics $\mathbf{x}^*(t)$ on the limit cycle Γ. For this theorem, we require three assumptions of regular conditions:

1. The functions $\varphi(\mathbf{x},t)$ and $\ln\omega(\mathbf{x},t)$ are smooth enough with respect to $\mathbf{x}$, and the derivatives uniformly converge on $\mathcal{D}$ as $t\to\infty$, in which the compact domain $\mathcal{D}$ is defined above.

2. The drift function $\mathbf{b}(\mathbf{x})$ and its Jacobian $\mathbf{A}(\mathbf{x})$ are continuous on $\mathcal{D}$.

3. For all $t\geq 0$, the covariance matrix $\boldsymbol{\Sigma}(t)$ in Lemma 6.2.4 is nonsingular, i.e, the Gaussian fluctuations of each direction is nonzero. Therefore, the inverse of Eq. (6.37) in Lemma 6.2.4 is well-defined: $[\boldsymbol{\Sigma}(t)]^{-1} = \nabla\nabla\varphi(\hat{\mathbf{x}}(t),t)$ for all $t\geq 0$.

Theorem 6.3.1. *Let $\mathbf{x}^*(t)$ be a deterministic trajectory with $\mathbf{x}^*(0)\in\Gamma$, $[\boldsymbol{\Sigma}^*(t)]^{-1} := \nabla\nabla\varphi(\mathbf{x}^*(t))$, and $\omega^*(t) := \omega(\mathbf{x}^*(t))$. For all $t>0$,*

$$\frac{\mathrm{d}\,[\boldsymbol{\Sigma}^*(t)]^{-1}}{\mathrm{d}t} = -[\boldsymbol{\Sigma}^*]^{-1}\mathbf{A}(\mathbf{x}^*) - \mathbf{A}(\mathbf{x}^*)^T[\boldsymbol{\Sigma}^*]^{-1} - 2[\boldsymbol{\Sigma}^*]^{-1}\mathbf{D}[\boldsymbol{\Sigma}^*]^{-1}, \tag{6.64}$$

$$\frac{\mathrm{d}\ln\omega^*(t)}{\mathrm{d}t} = -\nabla\cdot\mathbf{b}(\mathbf{x}^*) - \mathbf{D}[\boldsymbol{\Sigma}^*]^{-1} = -\nabla\cdot\boldsymbol{\gamma}(\mathbf{x}^*). \tag{6.65}$$

Furthermore, the smallest eigenvalues of the matrix $[\mathbf{\Sigma}^*(t)]^{-1}$ *is zero with the eigenvector tangential to* Γ *and the other eigenvalues are positive with the eigenvectors in the hyperplane perpendicular to* Γ.

Proof. In this proof, the norm $\|\cdot\|$ represents the supremum norm. By our setup of the diffusion process (6.50), the macroscopic deterministic trajectory $\hat{\mathbf{x}}(t)$, with $\hat{\mathbf{x}}(0) \in \mathcal{D}$, converges to the stable limit cycle Γ, so we have

$$\lim_{t\to\infty} \min_{\mathbf{x}^*\in\Gamma} \|\hat{\mathbf{x}}(t) - \mathbf{x}^*\| = 0. \tag{6.66}$$

By the assumption (1) and $\lim_{t\to\infty} \varphi(\mathbf{x}, t) = \varphi(\mathbf{x})$ in the setup, we have that

$$\lim_{t\to\infty} \nabla\nabla\varphi(\mathbf{x}, t) = \nabla\nabla\varphi(\mathbf{x}) \quad \text{uniformly on } \mathcal{D}. \tag{6.67}$$

Therefore, for any $\epsilon > 0$, there exists $T(\epsilon) > 0$ such that for every $t > T(\epsilon)$, there is a point $\mathbf{x}^*(t) \in \Gamma$ with its corresponding initial point $\mathbf{x}^*(0) \in \Gamma$ and

$$\|\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t) - \nabla\nabla\varphi(\mathbf{x}^*(t))\| = O(\epsilon), \tag{6.68}$$

in which we use the triangle inequality

$$\|\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t) - \nabla\nabla\varphi(\mathbf{x}^*(t))\| \le \|\nabla\nabla\varphi(\hat{\mathbf{x}}(t), t) - \nabla\nabla\varphi(\hat{\mathbf{x}}(t))\| + \|\nabla\nabla\varphi(\hat{\mathbf{x}}(t)) - \nabla\nabla\varphi(\mathbf{x}^*(t))\| \tag{6.69}$$

with Eq. (6.67) for the first term and Eq. (6.66) for the second term on the right side of the inequality. Apply $[\mathbf{\Sigma}^*(t)]^{-1} := \nabla\nabla\varphi(\mathbf{x}^*(t))$ defined in the theorem, and $[\mathbf{\Sigma}(t)]^{-1} = \nabla\nabla\varphi(\hat{\mathbf{x}}(t), t)$ given in Lemma 6.2.4 with the assumption (3), to Eq. (6.68), we further have that

$$\left\|[\mathbf{\Sigma}^*(t)]^{-1} - [\mathbf{\Sigma}(t)]^{-1}\right\| = O(\epsilon). \tag{6.70}$$

Following the same approach for the result (6.70) with the assumption (1) that the derivatives of φ uniformly converge on $\mathcal{D}$ and the assumption (2) that $\mathbf{b}$ is continuous on $\mathcal{D}$, we can also show that

$$\left\|\frac{\mathrm{d}\,[\mathbf{\Sigma}^*(t)]^{-1}}{\mathrm{d}t} - \frac{\mathrm{d}\,[\mathbf{\Sigma}(t)]^{-1}}{\mathrm{d}t}\right\| = O(\epsilon). \tag{6.71}$$

Furthermore, by the Lyapunov equation of $\mathbf{\Sigma}(t)$ obtained in Lemma 6.2.1, we have that

$$\frac{\mathrm{d}\left[\mathbf{\Sigma}(t)\right]^{-1}}{\mathrm{d}t} = \left[\mathbf{\Sigma}(t)\right]^{-1} \frac{\mathrm{d}\left[\mathbf{\Sigma}(t)\right]}{\mathrm{d}t} \left[\mathbf{\Sigma}(t)\right]^{-1} = -\left[\mathbf{\Sigma}\right]^{-1} \mathbf{A}(\hat{\mathbf{x}}) - \mathbf{A}(\hat{\mathbf{x}})^T \left[\mathbf{\Sigma}\right]^{-1} - 2\left[\mathbf{\Sigma}\right]^{-1} \mathbf{D}\left[\mathbf{\Sigma}\right]^{-1}. \tag{6.72}$$

By the assumption (2) that $\mathbf{A}$ is continuous on $\mathcal{D}$, combined with Eq. (6.70), Eq. (6.71) and Eq. (6.72), we can show that

$$\frac{\mathrm{d}\left[\mathbf{\Sigma}^*(t)\right]^{-1}}{\mathrm{d}t} = -\left[\mathbf{\Sigma}^*\right]^{-1} \mathbf{A}(\mathbf{x}^*) - \mathbf{A}(\mathbf{x}^*)^T \left[\mathbf{\Sigma}^*\right]^{-1} - 2\left[\mathbf{\Sigma}^*\right]^{-1} \mathbf{D}\left[\mathbf{\Sigma}^*\right]^{-1} + O(\epsilon), \tag{6.73}$$

in which we use the triangular inequality. Eq. (6.73) is true for any $\epsilon > 0$ and $t > T(\epsilon)$, and furthermore, the functions $\left[\mathbf{\Sigma}^*(t)\right]^{-1}$ and $\mathbf{A}(\mathbf{x}^*(t))$ are periodic by the definitions, so if Eq. (6.73) holds for $t > T(\epsilon)$, by the phase shift, it should hold for all $t > 0$. Then we can take $\epsilon \to 0$ to obtain

$$\frac{\mathrm{d}\left[\mathbf{\Sigma}^*(t)\right]^{-1}}{\mathrm{d}t} = -\left[\mathbf{\Sigma}^*\right]^{-1} \mathbf{A}(\mathbf{x}^*) - \mathbf{A}(\mathbf{x}^*)^T \left[\mathbf{\Sigma}^*\right]^{-1} - 2\left[\mathbf{\Sigma}^*\right]^{-1} \mathbf{D}\left[\mathbf{\Sigma}^*\right]^{-1}, \tag{6.74}$$

for all $t > 0$ with the initial condition $\left[\mathbf{\Sigma}^*(0)\right]^{-1} = \nabla\nabla\varphi(\mathbf{x}^*(0))$.

Next, we need to investigate the solution $\left[\mathbf{\Sigma}^*(t)\right]^{-1}$ given by the ODE (6.74). Let us introduce a coordinate transformation from the Cartesian coordinate to the curvilinear coordinate around the limit cycle by the change of basis

$$\mathbf{K}(t) = \mathbf{Q}(t)^{-1}[\mathbf{\Sigma}^*(t)]^{-1}\mathbf{Q}(t) \quad \text{with} \quad \mathbf{Q}(t) = \left[\mathbf{e}_1(t)\; \mathbf{e}_2(t)\; \cdots\; \mathbf{e}_n(t)\right], \tag{6.75}$$

in which $\mathbf{Q}(t)$ is a time-dependent orthonormal basis. In particular, $\mathbf{e}_1(t) = \mathbf{b}(\mathbf{x}^*)/\|\mathbf{b}(\mathbf{x}^*\|$ is the tangential unit vector on Γ and the span of the rest of vectors $\{\mathbf{e}_2(t)\; \cdots\; \mathbf{e}_n(t)\}$ represents the hyperplane perpendicular to Γ. For every fixed time t, $\mathbf{Q}(t)$ can be obtain by the *Gram–Schmidt process* and $\mathbf{Q}(t)$ is known as the *Frenet frame*. By the proposition (2), since $\nabla\nabla\varphi(\mathbf{x}^*)$ is always zero on the direction tangential to Γ, $\mathbf{e}_1(t)$ is in the nullspace of $\nabla\nabla\varphi(\mathbf{x}^*)$ for all t. With the fact $\left[\mathbf{\Sigma}^*(t)\right]^{-1} := \nabla\nabla\varphi(\mathbf{x}^*(t))$ is symmetric, for all $1 \le i \le n$ and $1 \le j \le n$, we have that

$$\left[\mathbf{K}(t)\right]_{i1} = \mathbf{e}_i(t)^T[\mathbf{\Sigma}^*(t)]^{-1}\mathbf{e}_1(t) \equiv 0 \quad \text{and} \quad \left[\mathbf{K}(t)\right]_{1j} = \mathbf{e}_1(t)^T[\mathbf{\Sigma}^*(t)]^{-1}\mathbf{e}_j(t) \equiv 0, \tag{6.76}$$

thus the matrix $\mathbf{K}(t)$ has both zero first column and zero first row. Therefore, we can define a submatrix $\tilde{\mathbf{K}}(t)$ by deleting the first column and the first row of $\mathbf{K}(t)$, equipped with Eq. (6.73), we will obtain a $n-1$ by $n-1$ system of differential equations

$$\frac{\mathrm{d}\tilde{\mathbf{K}}(t)}{\mathrm{d}t} = -\tilde{\mathbf{K}}(t)\left[\tilde{\mathbf{A}}(\mathbf{x}^*) - \tilde{\mathbf{S}}(t)\right] - \left[\tilde{\mathbf{A}}(\mathbf{x}^*) - \tilde{\mathbf{S}}(t)\right]^T \tilde{\mathbf{K}}(t) - 2\tilde{\mathbf{K}}(t)\tilde{\mathbf{D}}\tilde{\mathbf{K}}(t), \tag{6.77}$$

in which the symbol $\sim$ on top of each matrix represents the restriction of the original matrix in the hyperplane perpendicular to Γ and the additional term $\tilde{\mathbf{S}}(t)$, $\mathbf{S} = \mathbf{Q}^{-1}(t)\dot{\mathbf{Q}}(t)$, is from the coordinate transformation. We can identify that Eq. (6.77) is a *periodic Riccati differential equation*. Under the assumptions that Γ is a stable limit cycle and $\tilde{\mathbf{D}}$ is definite positive (the later follows from the definite positive $\mathbf{D}$ in the setup (6.50)), the solution of the periodic Riccati differential equation (6.77) has to be positive definite and periodic with the same period of the limit cycle. Mathematical analysis of this type of periodic Riccati differential equations can be found in the works [19, 146, 29, 221] and a comprehensive numerical analysis with several examples was provided in the work [126].

The proof for Eq. (6.65) of $\ln\omega^*(t)$ follows the proof for Eq. (6.64) of $[\boldsymbol{\Sigma}^*(t)]^{-1}$. Apply the same asymptotic analysis (Eq. (6.66) - Eq. (6.74)) to the dynamics of $\omega(\hat{\mathbf{x}}(t), t)$ in Eq. (6.63), we can show that $\ln\omega^*$ satisfies

$$\frac{\mathrm{d}\ln\omega^*(t)}{\mathrm{d}t} = \nabla\cdot\mathbf{b}(\mathbf{x}^*) - \mathbf{D}\nabla\nabla\varphi(\mathbf{x}^*) = -\nabla\cdot\boldsymbol{\gamma}(\mathbf{x}^*), \tag{6.78}$$

for all $t > 0$ with the initial condition $\ln\omega^*(0) = \ln\omega(\mathbf{x}^*(0))$.

□

To the best of our knowledge, the asymptotic analysis from Eq. (6.66) to Eq. (6.74) is the first work rigorously shows the Lyapunov differential equation (6.74) for the curvature of large deviation rate function near a deterministic trajectory from a transient state to an invariant set. Additionally, by analyzing the solution given by the Lyapunov differential equation with the coordinate transformation, we confirm that the solution $[\boldsymbol{\Sigma}^*(t)]^{-1}$ is consistent with the features of $\nabla\nabla\varphi(\mathbf{x})$ on Γ in the proposition (3) from the previous work. We will apply Theorem 6.3.1 to further obtain three characterizations of stochastic limit cycles: (i) probability flux near the cycle

(Sec. 6.3.2); (ii) two special features of $\boldsymbol{\gamma}$ on the limit cycle (Sec. 6.3.3); (iii) a local entropy balance equation on the cycle (Sec. 6.3.4).

6.3.2 Probability flux of near a limit cycle

Recall that the vector field $\boldsymbol{\gamma}$ is defined by the orthogonal decomposition of the drift $\mathbf{b}$ in Eq. (6.59) and Eq. (6.60), which characterizes the direction of circular motion of the deterministic dynamics $\hat{\mathbf{x}}' = \mathbf{b}(\hat{\mathbf{x}})$. In addition to the orthogonal decomposition of the drift $\mathbf{b}$, $\boldsymbol{\gamma}$ can be derived from the probability flux $\mathbf{J}[p_\epsilon(\mathbf{x},t)]$ defined in the FPE (6.52):

$$\boldsymbol{\gamma}_\epsilon(\mathbf{x},t) := \frac{\mathbf{J}[p_\epsilon(\mathbf{x},t)]}{p_\epsilon(\mathbf{x},t)} = \mathbf{b}(\mathbf{x}) - \epsilon\mathbf{D}\nabla\ln p_\epsilon(\mathbf{x},t), \tag{6.79}$$

and take $\epsilon \to 0$ before $t \to \infty$, we have that

$$\lim_{t\to\infty}\lim_{\epsilon\to 0}\boldsymbol{\gamma}_\epsilon(\mathbf{x},t) = \lim_{t\to\infty}\boldsymbol{\gamma}(\mathbf{x},t) = \boldsymbol{\gamma}(\mathbf{x}), \tag{6.80}$$

in which $\boldsymbol{\gamma}(\mathbf{x},t)$ is the same one defined in the HJE (6.53). For the stationary probability flux $\mathbf{J}[\pi_\epsilon(\mathbf{x})]$,

$$\boldsymbol{\gamma}_\epsilon(\mathbf{x}) := \frac{\mathbf{J}[\pi_\epsilon(\mathbf{x})]}{\pi_\epsilon(\mathbf{x})} = \mathbf{b}(\mathbf{x}) - \epsilon\mathbf{D}\nabla\ln \pi_\epsilon(\mathbf{x}), \tag{6.81}$$

which is corresponding to the reverse order of limits

$$\lim_{\epsilon\to 0}\lim_{t\to\infty}\boldsymbol{\gamma}_\epsilon(\mathbf{x},t) = \lim_{\epsilon\to 0}\boldsymbol{\gamma}_\epsilon(\mathbf{x}) = \boldsymbol{\gamma}(\mathbf{x}). \tag{6.82}$$

Note that $\boldsymbol{\gamma}_\epsilon(\mathbf{x})$ has been recognized as Onsager's thermodynamics force [157] or "local velocity" of the probability flux. Here we don't need to worry about the orders of limits if we just focus on the domain $\mathcal{D}$ based on our discussion in Sec. 6.3.1.

Recall that $\mathbf{A}(\mathbf{x})$ is the Jacobian matrix of $\mathbf{b}(\mathbf{x})$ and $[\boldsymbol{\Sigma}^*]^{-1} = \nabla\nabla\varphi(\mathbf{x}^*(t))$. By the above setup, we have the following theorem for $\boldsymbol{\gamma}_\epsilon(\mathbf{x})$ and the stationary probability flux $\mathbf{J}[\pi_\epsilon(\mathbf{x})]$ near Γ.

Theorem 6.3.2. *For* $\mathbf{x}^* \in \Gamma$ *and* $\|\mathbf{x}-\mathbf{x}^*\| = O(\sqrt{\epsilon})$,

$$\boldsymbol{\gamma}_\epsilon(\mathbf{x}) := \frac{\mathbf{J}[\pi_\epsilon(\mathbf{x})]}{\pi_\epsilon(\mathbf{x})} = \left[\mathbf{b}(\mathbf{x}^*) + \left(\mathbf{A}(\mathbf{x}^*) + \mathbf{D}\left[\boldsymbol{\Sigma}^*\right]^{-1}\right)(\mathbf{x}-\mathbf{x}^*)\right] + O(\epsilon), \tag{6.83}$$

where $\mathbf{D}\left[\boldsymbol{\Sigma}^*\right]^{-1}$ *is the Frobenius product, and* $\left[\boldsymbol{\Sigma}^*\right]^{-1}$ *satisfies the equation*

$$\frac{\mathrm{d}\left[\boldsymbol{\Sigma}^*(t)\right]^{-1}}{\mathrm{d}t} = -\left[\boldsymbol{\Sigma}^*\right]^{-1}\mathbf{A}(\mathbf{x}^*) - \mathbf{A}(\mathbf{x}^*)^T\left[\boldsymbol{\Sigma}^*\right]^{-1} - 2\left[\boldsymbol{\Sigma}^*\right]^{-1}\mathbf{D}\left[\boldsymbol{\Sigma}^*\right]^{-1}. \tag{6.84}$$

Proof. We first approximate $\nabla \ln \pi_\epsilon(\mathbf{x})$ near Γ, in which π_ϵ has the WKB expansion in Eq. (6.55). Let $\|\mathbf{x} - \mathbf{x}^*\| = O(\sqrt{\epsilon})$,

$$\begin{aligned}\nabla \ln \pi_\epsilon(\mathbf{x}) &= \nabla\left(\ln\omega(\mathbf{x}) - \frac{\varphi(\mathbf{x})}{\epsilon}\right) + O(\epsilon)\\ &= \nabla\left(\ln\omega(\mathbf{x}) - \frac{(\mathbf{x}-\mathbf{x}^*)^T\nabla\nabla\varphi(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)}{2\epsilon}\right) + O(1)\\ &= -\frac{\nabla\nabla\varphi(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)}{\epsilon} + O(1),\end{aligned} \tag{6.85}$$

where we use $\varphi(\mathbf{x}^*) \equiv 0$ and $\nabla\varphi(\mathbf{x}^*) \equiv 0$. Apply the approximation (6.85) to Eq. (6.81), we can further approximate $\boldsymbol{\gamma}_\epsilon(\mathbf{x})$ around Γ

$$\begin{aligned}\boldsymbol{\gamma}_\epsilon(\mathbf{x}) := \frac{\mathbf{J}[\pi_\epsilon(\mathbf{x})]}{\pi_\epsilon(\mathbf{x})} &= \mathbf{b}(\mathbf{x}) - \epsilon\mathbf{D}\nabla\ln\pi_\epsilon(\mathbf{x})\\ &= \mathbf{b}(\mathbf{x}^*) + \left(\mathbf{A}(\mathbf{x}^*) + \mathbf{D}\left[\boldsymbol{\Sigma}^*\right]^{-1}\right)(\mathbf{x}-\mathbf{x}^*) + O(\epsilon).\end{aligned} \tag{6.86}$$

The dynamics (6.84) is from Eq. (6.64) in Theorem 6.3.1.

□

Corollary 6.3.3. *In the particular case of linear dynamics* $\hat{\mathbf{x}}' = \mathbf{A}\hat{\mathbf{x}}$ *(*$\mathbf{A}$ *is a constant* $n \times n$ *matrix) with a stable fixed point* $\mathbf{x}^* = 0$*, the stationary probability flux* $\mathbf{J}$ *has the formula [160]*

$$\mathbf{J}[\pi_\epsilon(\mathbf{x})] = \pi_\epsilon^{-1}(\mathbf{x})\left[\left(\mathbf{A} + \mathbf{D}\boldsymbol{\Sigma}^{-1}\right)\mathbf{x}\right], \tag{6.87}$$

where $\boldsymbol{\Sigma}^{-1}$ *satisfies* $\boldsymbol{\Sigma}^{-1}\mathbf{A} + \mathbf{A}^T\boldsymbol{\Sigma}^{-1} + 2\boldsymbol{\Sigma}^{-1}\mathbf{D}\boldsymbol{\Sigma}^{-1} = 0$*. To compare Eq.* (6.83) *with Eq.* (6.87)*, Theorem 6.3.2 can be regarded as a local linear approximation of the stationary probability flux near each point* $\mathbf{x}^*$ *on the limit cycle. This new result extends the case from a fix point to an invariant set.*

6.3.3 Two features of $\boldsymbol{\gamma}$ on the limit cycle

By Theorem 6.3.2, we have an approximate the probability flux near Γ, and furthermore, by the stationary FPE (6.52), there is another important property of the probability flux

$$\nabla \cdot \mathbf{J}[\pi_\epsilon(\mathbf{x})] = 0, \quad \text{for all } \mathbf{x} \in \mathbb{R}^n. \tag{6.88}$$

Since this property holds in the whole space, we can apply it to an arbitrary neighborhood of the limit cycle. Having this property, we will obtain two special features of $\boldsymbol{\gamma}$ on the limit cycle in this section.

The derivation of the system of equations (6.59) - (6.61) in the work [171] was by plugging the WKB ansatz (6.55) into the stationary FPE and equating like-order terms of

$$\nabla \cdot (\pi_\epsilon(\mathbf{x})\boldsymbol{\gamma}_\epsilon(\mathbf{x})) = \nabla \cdot \mathbf{J}[\pi_\epsilon(\mathbf{x})] = 0, \quad \text{for all } \mathbf{x} \in \mathbb{R}^n. \tag{6.89}$$

Applying Eq. (6.61) in the system of equations to Γ, we obtain the first feature of $\boldsymbol{\gamma}$ on Γ

$$\nabla \cdot (\omega(\mathbf{x})\boldsymbol{\gamma}(\mathbf{x})) = -\nabla\varphi(\mathbf{x}) \cdot \mathbf{D}\nabla\omega(\mathbf{x}) = 0 \quad \text{for all } \mathbf{x} \in \Gamma, \tag{6.90}$$

where we use $\nabla\varphi(\mathbf{x}) \equiv 0$ for $\mathbf{x} \in \Gamma$. We can recognize that the divergence-free stationary probability flux in $\mathbb{R}^n$ is the key to get Eq. (6.89) so that we can further obtain the divergence-free $\omega(\mathbf{x})\boldsymbol{\gamma}(\mathbf{x})$ on the limit cycle in Eq. (6.90). With a similar approach, not only the divergence of $\omega(\mathbf{x})\boldsymbol{\gamma}(\mathbf{x})$, we can also obtain the second feature, $||\omega(\mathbf{x})\boldsymbol{\gamma}(\mathbf{x})||$, on the limit cycle via the following theorem:

Theorem 6.3.4. *Let $1/v(\mathbf{x})$ be the product of the nonzero eigenvalues of the matrix $\nabla\nabla\varphi(\mathbf{x})$. Then*

$$\sqrt{v(\mathbf{x})} \times ||\omega(\mathbf{x})\boldsymbol{\gamma}(\mathbf{x})|| \tag{6.91}$$

is constant on the limit cycle Γ. Furthermore, let $g_\epsilon(\mathbf{x})$ be the marginal density of $\pi_\epsilon(\mathbf{x})$ on the limit cycle Γ, then for $\mathbf{x} \in \Gamma$,

$$g_\epsilon(\mathbf{x}) = \frac{\omega(\mathbf{x})\sqrt{v(\mathbf{x})}}{\int_\Gamma \omega(\mathbf{y})\sqrt{v(\mathbf{y})}d\mathbf{y}} + O(\epsilon), \tag{6.92}$$

and there exists a constant K such that $g_\epsilon(\mathbf{x})||\boldsymbol{\gamma}(\mathbf{x})|| = K + O(\epsilon)$.

From the preceding discussion in Sec. 6.3.1, since $\varphi(\mathbf{x})$ is constant on Γ, the eigenvector of $\nabla\nabla\varphi(\mathbf{x})$ corresponding to the only one zero eigenvalue is tangential to Γ. Therefore, $v(\mathbf{x})$ in Theorem 6.3.4 defined on Γ represents the scaled variance in the hyperplane perpendicular to Γ (Recall that this hyperplane is defined by the span of the vectors $\mathbf{e}_2, \cdots, \mathbf{e}_n$ in the coordinate transformation (6.75)). By Eq. (6.90), we have that $\omega(\mathbf{x})\boldsymbol{\gamma}(\mathbf{x})$ is divergence-free on the limit cycle. By Theorem 6.3.4, we further have that $||\omega(\mathbf{x})\boldsymbol{\gamma}(\mathbf{x})||$ is reciprocal to the scaled standard deviation perpendicular to the limit cycle. The later was mentioned in the previous work [200]. In the present work, we provide a mathematical proof in Appendix 6.6.2. The idea of proof is by using the Gauss's theorem for a tube around the limit cycle. Since the Gauss's theorem can only be applied to a small but finite tube, the divergence for the Gauss's theorem we use in the proof is $\nabla \cdot (\boldsymbol{\gamma}_\epsilon(\mathbf{x})\pi_\epsilon(\mathbf{x})) = \nabla \cdot \mathbf{J}[\pi_\epsilon(\mathbf{x})] = 0$, which holds for an arbitrary neighborhood of Γ.

6.3.4 A local entropy balance equation on the limit cycle

On the limit cycle, we have derived the local Gaussian fluctuations of dynamics represented by the covariance $\boldsymbol{\Sigma}^*(t)$ which follows the periodic Lyanupov equation (6.64) in Theorem 6.3.1. In general, the entropy of a Gaussian distribution p with a covariance $\boldsymbol{\Sigma}$ is

$$S = -\int p(\mathbf{x}) \ln p(\mathbf{x}) \mathrm{d}\mathbf{x} = \frac{1}{2} \ln \left[2\pi \det(\boldsymbol{\Sigma})\right]. \tag{6.93}$$

Therefore, in nonlinear stochastic systems, the "local" entropy (denoted by S_l) due to the local Gaussian fluctuations has the rate defined by

$$\frac{\mathrm{d}S_l(t)}{\mathrm{d}t} := \frac{1}{2}\frac{\mathrm{d}\ln \det(\boldsymbol{\Sigma}^*(t))}{\mathrm{d}t}. \tag{6.94}$$

By the property that the determinant of a matrix equals to the product of its eigenvalues, the local rate of entropy change (6.94) has an equivalent definition,

$$\frac{\mathrm{d}S_l(t)}{\mathrm{d}t} := -\frac{1}{2}\frac{\mathrm{d}\left(\sum_{k=1}^n \ln \lambda_k^*(t)\right)}{\mathrm{d}t}, \tag{6.95}$$

where $\lambda_k^*(t),\ 1 \le k \le n$ are the eigenvalues of $[\boldsymbol{\Sigma}^*(t)]^{-1}$. Note that $1/v(\mathbf{x}^*(t))$ defined in Theorem 6.3.4 equals to the product of all nonzero eigenvalues $\lambda_2^*(t) \cdots \lambda_n^*(t)$ of the matrix $[\boldsymbol{\Sigma}^*(t)]^{-1}$.

By the above setup, we have the following theorem of a *local entropy balance equation* on the limit cycle Γ with three equivalent expressions.

Theorem 6.3.5. *For $\mathbf{x}^*(t) \in \Gamma$, by the definition* (6.94) *of the local rate of entropy change, there exists a local entropy balance equation with three equivalent expressions,*

$$\frac{\mathrm{d}S_l(t)}{\mathrm{d}t} = \nabla \cdot \boldsymbol{\gamma}(\mathbf{x}^*(t)), \tag{6.96}$$

$$= -\frac{\mathrm{d}\ln\omega(\mathbf{x}^*(t))}{\mathrm{d}t}, \tag{6.97}$$

$$= \frac{\mathrm{d}\ln||\boldsymbol{\gamma}(\mathbf{x}^*(t))||}{\mathrm{d}t} + \frac{1}{2}\frac{\mathrm{d}\ln v(\mathbf{x}^*(t))}{\mathrm{d}t}. \tag{6.98}$$

Proof. By Eq. (6.94), with the dynamics of $[\boldsymbol{\Sigma}^*(t)]^{-1}$ (6.64), we can obtain

$$\frac{\mathrm{d}S_l(t)}{\mathrm{d}t} = \nabla \cdot \mathbf{b}(\mathbf{x}^*) + \mathbf{D}\,[\boldsymbol{\Sigma}^*]^{-1} = \nabla \cdot \boldsymbol{\gamma}(\mathbf{x}^*), \tag{6.99}$$

where $\mathbf{D}\,[\boldsymbol{\Sigma}^*]^{-1}$ is the Frobenius product of the matrix $\mathbf{D}$ and the matrix $[\boldsymbol{\Sigma}^*]^{-1}$. Furthermore, by Eq. (6.65) for the prefactor ω, we can link Eq. (6.99) to the dynamics of ω,

$$\frac{\mathrm{d}S_l(t)}{\mathrm{d}t} = -\frac{\mathrm{d}\ln\omega(\mathbf{x}^*(t))}{\mathrm{d}t}. \tag{6.100}$$

So far, we have proved the first two expressions (6.96) and (6.97). The following proof is for the third expression (6.98): By Theorem 6.3.1, we know the smallest eigenvalue $\lambda_1^*(t) \equiv 0$ with its eigenvector tangential to the limit cycle. Therefore, the first term on the right side of Eq. (6.95) requires a further analysis since

$$\frac{\mathrm{d}\ln\lambda_1^*(t)}{\mathrm{d}t} = \frac{1}{\lambda_1^*(t)}\frac{\mathrm{d}\lambda_1^*(t)}{\mathrm{d}t} = \infty \times 0. \tag{6.101}$$

To find an explicit formula of (6.101), we can use

$$\frac{\mathrm{d}\left(\boldsymbol{\gamma}^T(\mathbf{x}^*(t))\,[\boldsymbol{\Sigma}^*(t)]^{-1}\,\boldsymbol{\gamma}(\mathbf{x}^*(t))\right)}{\mathrm{d}t} = 0, \quad \text{for all } t > 0, \tag{6.102}$$

since $[\boldsymbol{\Sigma}^*(t)]^{-1}\,\boldsymbol{\gamma}(\mathbf{x}^*(t)) \equiv \mathbf{0}$ by Theorem 6.3.1. By rearranging Eq. (6.102), we find a formula of Eq. (6.101),

$$\frac{\mathrm{d}\ln\lambda_1^*(t)}{\mathrm{d}t} = -2\frac{\mathrm{d}\ln\|\boldsymbol{\gamma}(\mathbf{x}^*(t))\|}{\mathrm{d}t}, \tag{6.103}$$

where we use that $\lambda_1^*(t)$ is the eigenvalue of $[\mathbf{\Sigma}^*(t)]^{-1}$ with respect to the eigenvector $\boldsymbol{\gamma}(\mathbf{x}^*(t))$. Therefore, by Eq. (6.95) and Eq. (6.103), and with the definition of $1/v(\mathbf{x}^*(t)) := \prod_{k=2}^n \lambda_k^*(t)$, the local rate of entropy change on Γ has another expression

$$\frac{\mathrm{d}S_l(t)}{\mathrm{d}t} = \frac{\mathrm{d}\ln||\boldsymbol{\gamma}(\mathbf{x}^*(t))||}{\mathrm{d}t} + \frac{1}{2}\frac{\mathrm{d}\ln v(\mathbf{x}^*(t))}{\mathrm{d}t}. \tag{6.104}$$

□

Each expression has a clear physical meaning:

1. The first expression (6.96): The divergence of a vector field characterizes the volume change of the flow following this vector field. Therefore, the local entropy change can be considered as a consequence of volume-expanding (entropy-increasing) or volume-contracting (entropy-decreasing) of the circular flow $\mathbf{x}^*(t)' = \boldsymbol{\gamma}(\mathbf{x}^*)$. This expression of the entropy balance is corresponding to the *microscopic entropy production rate* given by a large number particle system without noise [41, 187].

2. The second expression (6.97): Let us compare the rate of free energy change [73] (the free energy is defined by the relative entropy $F(t) = \int_{\mathbb{R}^n} p(\mathbf{x},t)\log\left(\frac{p(\mathbf{x},t)}{\pi(\mathbf{x})}\right)\mathrm{d}\mathbf{x}$) with the local rate of entropy change on the limit cycle Γ:

$$\frac{\mathrm{d}F(t)}{\mathrm{d}t} = \frac{\mathrm{d}\varphi(\mathbf{x}^*(t))}{\mathrm{d}t} \equiv 0, \tag{6.105}$$
$$\frac{\mathrm{d}S_l(t)}{\mathrm{d}t} = -\frac{\mathrm{d}\ln\omega(\mathbf{x}^*(t))}{\mathrm{d}t}. \tag{6.106}$$

   The former follows the change of the large-deviation rate function $\varphi(\mathbf{x})$ on the deterministic trajectory, which is always zero on the limit cycle due to constant $\varphi(\mathbf{x}^*(t))$; The later follows the change of $-\ln\omega(\mathbf{x}^*(t))$, where the prefactor $\omega(\mathbf{x})$ is known as "degeneracy" in the classical statistical mechanical terminology [171], which is not constant on the limit cycle in general.

3. The third expression (6.98): The local entropy balance equation can be decomposed into two parts

$$\frac{\mathrm{d}S_l(t)}{\mathrm{d}t} = \underbrace{\frac{\mathrm{d}\ln||\boldsymbol{\gamma}(\mathbf{x}^*(t))||}{\mathrm{d}t}}_{\text{dissipative part}} + \underbrace{\frac{1}{2}\frac{\mathrm{d}\ln v(\mathbf{x}^*(t))}{\mathrm{d}t}}_{\text{fluctuation part}}. \tag{6.107}$$

   The first part is yielded by the change of speed on the limit cycle, which is determined from the deterministic path of the dissipative dynamics $\mathbf{x}^*(t)' = \boldsymbol{\gamma}(\mathbf{x}^*)$; The second part is constituted by the change of the Gaussian fluctuations perpendicular to Γ. The *fluctuation-dissipation theory of nonequilibrium systems* by Keizer [107] elucidated the relation between the fluctuations of a time-inhomogeneous Gaussian process and the associated dissipative deterministic path. Following this theory, Eq. (6.107) can be regarded as a *fluctuation-dissipation decomposition* of the local entropy balance equation on the limit cycle.

Corollary 6.3.6. *By integrating the second expression* (6.97)*, when the system reaches its steady state, we have an equation of local entropy near the limit cycle,*

$$S_l(t) = -\ln\omega(\mathbf{x}^*(t)) + C, \tag{6.108}$$

for some constant C. By the equation of entropy (6.108)*, we know that in the long run, the entropy of system measured near the limit cycle in the scope of the CLT should be* periodic *with the same period of the cycle. On the other hand, the global entropy in the total system has to be constant in the long run because of the existence of stationary distribution.*

Corollary 6.3.7. *By the equivalence of the expressions* (6.97) *and* (6.98) *in Theorem 6.3.5, we have an alternative proof for the constant $\sqrt{v(\mathbf{x})} \times ||\omega(\mathbf{x})\boldsymbol{\gamma}(\mathbf{x})||$ on the limit cycle Γ in Theorem 6.3.4.*

6.4 Related issue: the scaling hypothesis of diffusion processes

As the success of the well-established scaling hypothesis in the continuous-time non-negative integer valued Markov population process $\mathbf{n}_V(t)$ (it has a law of large number as the system's size

$V \to \infty : V^{-1}\mathbf{n}_V(t) \to \mathbf{c}(t)$, the concentration of all the species [116]), we shall justify the origin of ϵ in (6.50) with physical interpretations more than just a mathematical tool.

Let us begin with a diffusion process $\mathbf{Y}(\tau) \in \mathbb{R}^n$ satisfied the following SDE

$$\mathrm{d}\mathbf{Y}(\tau) = \mathbf{g}(\mathbf{Y})\mathrm{d}\tau + [2\mathbf{D}(\mathbf{Y})]^{\frac{1}{2}}\mathrm{d}\mathbf{B}(\tau), \tag{6.109}$$

where $\boldsymbol{\gamma} : \mathbb{R}^n \to \mathbb{R}^n$ stands for the drift of the process, $\mathbf{D} : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ is the diffusion matrix, and $\mathbf{B}(\tau)$ is the standard n-dimensional Brownian motion. Through choosing different scales, $\mathbf{X} = \mathbf{Y}/\alpha$, $t = \tau/\beta$, the SDE (6.109) can be rescaled as

$$\mathrm{d}\mathbf{X}(t) = \frac{\beta}{\alpha}\mathbf{g}(\alpha\mathbf{X})\mathrm{d}t + \frac{\sqrt{\beta}}{\alpha}[2\mathbf{D}(\alpha\mathbf{X})]^{\frac{1}{2}}\mathrm{d}\mathbf{B}(t). \tag{6.110}$$

We assume a *space-time structure* $\beta = \xi(\alpha)$ by a function $\xi : \mathbb{R} \to \mathbb{R}$, and define a small parameter ϵ

$$\epsilon := \xi(\alpha)/\alpha^2 \tag{6.111}$$

with an implicit solution $\alpha^*(\epsilon)$ of Eq. (6.111). Under this framework, the scaled SDE (6.110) becomes a sequence of SDEs parameterized by ϵ

$$\mathrm{d}\mathbf{X}(t) = \mathbf{b}_\epsilon(\mathbf{X})\mathrm{d}t + [2\epsilon\mathbf{D}_\epsilon(\mathbf{X})]^{\frac{1}{2}}\mathrm{d}\mathbf{B}(t), \tag{6.112}$$

where

$$\mathbf{b}_\epsilon(\mathbf{x}) := \frac{\xi(\alpha^*(\epsilon))}{\alpha^*(\epsilon)}\boldsymbol{\gamma}\left(\alpha^*(\epsilon)\mathbf{x}\right) \quad \text{and} \quad \mathbf{D}_\epsilon(\mathbf{x}) := \mathbf{D}\left(\alpha^*(\epsilon)\mathbf{x}\right). \tag{6.113}$$

In order to observe an emergent phenomenon as $\epsilon \to 0$, we require certain conditions (i) $\lim_{\epsilon\to 0}\mathbf{b}_\epsilon(\mathbf{x}) = \mathbf{b}(\mathbf{x})$ exists. (ii) $\lim_{\epsilon\to 0}\mathbf{D}_\epsilon(\mathbf{x}) = \mathbf{D}(\mathbf{x})$ exists. (iii) The convergence of random processes solved of the SDEs (6.112) in different modes exists [65]. Then the limit gives us an emergent deterministic dynamics $\mathrm{d}\mathbf{x}(t)/\mathrm{d}t = \mathbf{b}(\mathbf{x})$ as $\epsilon \to 0$.

In connection to classical overdamped mechanical motions in a viscous fluid, Eq. (6.112) is widely called a Langevin equation, and in this case the small parameter ϵ has been identified as related to temperature of the system as well as the "scale" under which the mechanical motion is being observed [199]. In reality, the limit of ϵ being zero should be interpreted as particle motions in a "continuous medium at finite temperature" rather than "temperature asymptotic to zero". We follow this physical intuition and so called *scaling hypothesis* [103] for the origin of ϵ. The following discussion offers an insight into the connection between our scaling hypothesis for diffusion processes and the scaling hypothesis for statistical physics of fields.

The space-time structure defined by the function ξ is rather general. To illustrate our hypothesis, we focus on a specific space-time structure $\beta = \alpha^k$ and thus the small parameter is defined as $\epsilon := \alpha^{k-2}$. In this example, when $k < 2$, deterministic dynamics emerges at the macroscopic scale ($\alpha \to \infty$); when $k > 2$, deterministic dynamics emerges at the microscopic scale ($\alpha \to 0$). The choice of k depends on the property of the underlying drift function $\boldsymbol{\gamma}$ and the diffusion function $\mathbf{D}$. Given $\boldsymbol{\gamma}(\mathbf{x}) = c\mathbf{x}^n$, c is a constant, and $\mathbf{D}$ is a constant matrix, then the sequence of SDEs (6.112) becomes

$$\mathrm{d}\mathbf{X}(t) = \epsilon^{\frac{k-1+n}{k-1}} c\mathbf{X}^n \mathrm{d}t + [2\epsilon\mathbf{D}]^{\frac{1}{2}} \mathrm{d}\mathbf{B}(t). \tag{6.114}$$

In order to fulfill the conditions of convergence, in this example, the order of space-time structure k must be determined by the order of the underlying drift function n, i.e., $k = 1 - n$. Hence, the scaled drift function becomes ϵ-independent while the diffusion term is asymptotic to zero as $\epsilon \to 0$, which gives rise to an emergent deterministic dynamics $\mathrm{d}\mathbf{x}(t) = c\mathbf{x}^n\mathrm{d}t$. In other words, as a scientific theory, when we are able to observe deterministic dynamics $\mathrm{d}\mathbf{x}(t) = c\mathbf{x}^n\mathrm{d}t$ in a "macroscopic" experiment, with an underlying stochastic dynamics having the drift function $\boldsymbol{\gamma}(\mathbf{x}) = c\mathbf{x}^n$, $n > -1$, this experiment must be running by the right space-time structure $\beta = \alpha^{1-n}$.

In addition to the scale of space, by the space-time relation $\beta = \alpha^{1-n}$, for the macroscopic emergent deterministic dynamics ($\alpha \to \infty$), there is a corresponding scale of time for emergent laws: As $-1 < n < 1$, the deterministic dynamics emerges in a long-time limit; As $n > 1$, it emerges in a short-time limit. So emergent dynamics could be observed at different combinations of space-time scales, which are determined by n. This scaling exponent is given by the drift function $\boldsymbol{\gamma}$ of the underlying diffusion process. Therefore, in our scaling hypothesis for diffusion processes, experimental observation of a power law for the space-time structure is determined by the underlying physics. So we name it *scaling hypothesis*, which upholds the principle of scaling hypothesis for statistical physics of fields [103].

Here we want to introduce two types of celebrated theories which inspired our scaling hypothesis and point out what the new results we can provide beyond those theories:

1. As we mentioned in Sec. 6.1.2, the sequence of SDEs (6.112) has been carefully studied in the text *random perturbations of dynamical systems* by Freidlin and Wentzell [65]. Our scaling hypothesis gives ϵ a physical meaning which was unclear.

2. The Kurtz's first theorem [116] showed that the ODE model is an emergent model under the infinite volume limit of the discrete Markov chain model; And the Kurtz's second theorem [115] about the CLT for Markov chains is a generalization of a simple random walk for Donsker's invariance principle [43]. The main distinction between our scaling hypothesis and the scaling used in the Kurtz's theorems is that the former is for a sequence of *scaled stochastic differential equations* but the later is for a sequence of the sum over a *scaled Markov chain*.

6.5 Discussions and applications

In Sec. 6.2, we provided two preliminaries, Eq. (6.12) in Lemma 6.2.1 and Eq. (6.37) in Lemma 6.2.4: The former gives us the dynamics of the covariance of a time-inhomogeneous Gaussian process and the later characterizes the local curvature of the time-dependent large deviation function near it infimum. They both have a nice property that the existence of stationary probability is not required, so it helps us understand transient behaviors of the systems whose stationary probability may not exist. For example, if a system has unstable macroscopic deterministic dynamics, we can still compute its transient local Gaussian fluctuations and curvature of the rate function near the deterministic trajectory.

In Sec. 6.3, by asymptotic analysis, we characterized the dynamics near a stable limit cycle, and we found that the prefactor ω in the WKB ansatz plays an important role, which can be seen in Theorem 6.3.4 and Theorem 6.3.5. In contradistinction to the well-established theories [57, 62] of the HJE (6.53) for the large deviation rate function φ, to the best of our knowledge, sophisticated mathematical analysis of the PDE (6.63) for ω might be missing and it is worthy of attention in the future. For applications, the local entropy balance equation in Theorem 6.3.5 can help us seek a better understanding of thermodynamic behaviors of stochastic biological oscillators, e.g., (i) mammalian cell cycles under external noises [124], (ii) a modified Morris–Lecar conductance-based model of a neuron driven by extrinsic noise [21], and (iii) Rosenzweig-MacArthur model for predator-prey interactions with the effect of stochasticity [135].

In Sec. 6.4, the scaling hypothesis as a scientific theory, it allows us to apply the treatment to mathematical models of the complex systems whose "noise" does not have a clear origin as classical overdamped mechanical motions in a viscous fluid described by the Langevin equation. For example, the hypothesis could be used to the models of mechanical motions in biology with noise due to coarse graining. In addition to the justification of small parameter ϵ itself, this hypothesis may give us a clarification of the origin of ϵ-dependent drift function $\mathbf{b}$. It is known that there are two types of integrals for SDEs:

$$\mathrm{d}\mathbf{X}(t) = \mathbf{b}_I(\mathbf{X})\mathrm{d}t + [2\epsilon\mathbf{D}(\mathbf{X})]^{\frac{1}{2}}\mathrm{d}\mathbf{B}(t) \quad \text{(Itô interpretation)}, \tag{6.115}$$

$$\mathrm{d}\mathbf{X}(t) = \mathbf{b}_S(\mathbf{X})\mathrm{d}t + [2\epsilon\mathbf{D}(\mathbf{X})]^{\frac{1}{2}} \circ \mathrm{d}\mathbf{B}(t) \quad \text{(Stratonovich interpretation)}. \tag{6.116}$$

The former is commonly used in mathematical analysis and financial mathematics and the later is mostly applied in physics and engineering. Note that

$$\mathbf{b}_I(\mathbf{x}) = \mathbf{b}_S(\mathbf{x}) + \epsilon\nabla\cdot\mathbf{D}(\mathbf{x}). \tag{6.117}$$

Follow the scaling hypothesis, the existence of the extra ϵ-order term in Eq. (6.117) could be a corollary of the existence of higher-order terms in the underlying drift function before scaling. This hypothesis provides us a link between the two types of integral. It might help us to unravel the mystery of Itô - Stratonovich dilemma [199, 69, 4] in the future.

6.6 Appendix

6.6.1 Proofs of Lemma 6.2.2 and Lemma 6.2.3

Proof of Lemma 6.2.2

Proof. For the special case of $\mathbf{x} \in \mathbb{R}$, $\Theta = 3$, $\Lambda = 15$, and $\boldsymbol{\Xi} = [h''(\mathbf{x}^*)]^{-1}$. Then Eq. (6.29)

$$\eta(\mathbf{x}^*) = \frac{f''(\mathbf{x}^*)}{2h''(\mathbf{x})} - \left[\frac{f'(\mathbf{x}^*)h'''(\mathbf{x}^*)}{2[h''(\mathbf{x}^*)]^2} + \frac{f(\mathbf{x}^*)h''''(\mathbf{x}^*)}{8[h''(\mathbf{x}^*)]^2}\right] + \frac{5f(\mathbf{x}^*)[h'''(\mathbf{x}^*)]^2}{24[h'''(\mathbf{x}^*)]^3}.$$

This result can be found on P. 273 of [13], Eq. (6.4.35).

For the general case,

$$
\begin{aligned}
& \int_{\mathbb{R}^n} f(\mathbf{x}) e^{-\frac{h(\mathbf{x})}{\epsilon}} \mathrm{d}\mathbf{x} \\
= & \int_{\mathbb{R}^n} \left[f(\mathbf{x}^*) + (\mathbf{x}-\mathbf{x}^*) \cdot \nabla f(\mathbf{x}^*) + \frac{(\mathbf{x}-\mathbf{x}^*)^T \nabla\nabla f(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)}{2} + \dots \right] \\
\times & \exp\left[-\frac{h(\mathbf{x}^*)}{\epsilon} - \frac{(\mathbf{x}-\mathbf{x}^*)_i h''_{ij}(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)_j}{2\epsilon} - \frac{h'''_{ijk}(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)_i(\mathbf{x}-\mathbf{x}^*)_j(\mathbf{x}-\mathbf{x}^*)_k}{6\epsilon} \right. \\
- & \left. \frac{h''''_{ijk\ell}(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)_i(\mathbf{x}-\mathbf{x}^*)_j(\mathbf{x}-\mathbf{x}^*)_k(\mathbf{x}-\mathbf{x}^*)_\ell}{24\epsilon} + \cdots \right] \mathrm{d}\mathbf{x} \\
= & e^{-\frac{h(\mathbf{x}^*)}{\epsilon}} \int_{\mathbb{R}^N} \left[f(\mathbf{x}^*) + (\mathbf{x}-\mathbf{x}^*) \cdot \nabla f(\mathbf{x}^*) + \frac{(\mathbf{x}-\mathbf{x}^*)^T \nabla\nabla f(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)}{2} + \dots \right] \\
\times & \left[1 - \frac{h'''_{ijk}(\mathbf{x}^*)}{6\epsilon}(\mathbf{x}-\mathbf{x}^*)_i(\mathbf{x}-\mathbf{x}^*)_j(\mathbf{x}-\mathbf{x}^*)_k - \frac{h''''_{ijk\ell}(\mathbf{x}^*)}{24\epsilon}(\mathbf{x}-\mathbf{x}^*)_i(\mathbf{x}-\mathbf{x}^*)_j \right. \\
\times & \left. (\mathbf{x}-\mathbf{x}^*)_k(\mathbf{x}-\mathbf{x}^*)_\ell + \frac{\left[h'''_{ijk}(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)_i(\mathbf{x}-\mathbf{x}^*)_j(\mathbf{x}-\mathbf{x}^*)_k\right]^2}{72\epsilon^2} + \cdots \right] \\
\times & e^{-\frac{(\mathbf{x}-\mathbf{x}^*)^T \nabla\nabla h(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)}{2\epsilon}} \mathrm{d}\mathbf{x} \\
= & e^{-\frac{h(\mathbf{x}^*)}{\epsilon}} \int_{\mathbb{R}^n} e^{-\frac{(\mathbf{x}-\mathbf{x}^*)^T \nabla\nabla h(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)}{2\epsilon}} \left[f(\mathbf{x}^*) + (\mathbf{x}-\mathbf{x}^*) \cdot \nabla f(\mathbf{x}^*) \right. \\
+ & \frac{(\mathbf{x}-\mathbf{x}^*)^T \nabla\nabla f(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)}{2} - \frac{f(\mathbf{x}^*) h'''_{ijk}(\mathbf{x}^*)}{6\epsilon}(\mathbf{x}-\mathbf{x}^*)_i(\mathbf{x}-\mathbf{x}^*)_j(\mathbf{x}-\mathbf{x}^*)_k \\
- & \left(\frac{f'_i(\mathbf{x}^*) h'''_{jk\ell}(\mathbf{x}^*)}{6\epsilon} + \frac{f(\mathbf{x}^*) h''''_{ijk\ell}(\mathbf{x}^*)}{24\epsilon} \right) (\mathbf{x}-\mathbf{x}^*)_i(\mathbf{x}-\mathbf{x}^*)_j(\mathbf{x}-\mathbf{x}^*)_k(\mathbf{x}-\mathbf{x}^*)_\ell \\
+ & \left. \frac{f(\mathbf{x}^*)[h'''_{ijk}(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)_i(\mathbf{x}-\mathbf{x}^*)_j(\mathbf{x}-\mathbf{x}^*)_k]^2}{72\epsilon^2} + \cdots \right] \mathrm{d}\mathbf{x} \qquad (6.118) \\
= & \sqrt{\frac{(2\pi\epsilon)^N}{\det\left[\nabla\nabla h(x^*)\right]}} e^{-\frac{h(\mathbf{x}^*)}{\epsilon}} \left\{ f(\mathbf{x}^*) + \frac{\epsilon f''_{ij}(\mathbf{x}^*)}{2} \Xi_{ij}(\mathbf{x}^*) \right. \\
- & \epsilon \left[\frac{f'_i(\mathbf{x}^*) h'''_{jk\ell}(\mathbf{x}^*)}{6} + \frac{f(\mathbf{x}^*) h''''_{ijk\ell}(x^*)}{24} \right] \Xi^{\frac{1}{2}}_{i\mu} \Xi^{\frac{1}{2}}_{j\nu} \Xi^{\frac{1}{2}}_{k\rho} \Xi^{\frac{1}{2}}_{\ell\kappa} \Theta_{\mu\nu\rho\kappa} \\
+ & \left. \epsilon \left(\frac{f(\mathbf{x}^*)[h'''_{ijk}(\mathbf{x}^*)]^2}{72} \right) \Xi^{-\frac{1}{2}}_{i\mu} \Xi^{-\frac{1}{2}}_{i\mu'} \Xi^{-\frac{1}{2}}_{j\nu} \Xi^{-\frac{1}{2}}_{j\nu'} \Xi^{-\frac{1}{2}}_{k\rho} \Xi^{-\frac{1}{2}}_{k\rho'} \Lambda_{\mu\mu'\nu\nu'\rho\rho'} + \cdots \right\}.
\end{aligned}
$$

A multivariate normal distribution with covariance matrix Ξ, which is positive definite thus

$\mathbf{\Xi} = \mathbf{\Xi}^{\frac{1}{2}}\mathbf{\Xi}^{\frac{T}{2}}$ [20], has

$$\begin{aligned}
&\frac{1}{\left[\left(2\pi\epsilon\right)^N \det\left(\mathbf{\Xi}\right)\right]^{\frac{1}{2}}} \int_{\mathbb{R}^n} f''_{ij}(\mathbf{0})x_i x_j \exp\left[-\frac{1}{2\epsilon}\mathbf{x}^T\mathbf{\Xi}^{-1}\mathbf{x}\right] \mathrm{d}\mathbf{x} \\
= \quad & \frac{\epsilon\Xi^{\frac{1}{2}}_{i\nu}\Xi^{\frac{1}{2}}_{j\mu}f''_{ij}(\mathbf{0})}{(2\pi)^{N/2}} \int_{\mathbb{R}^n} y_\nu y_\mu \exp\left[-\frac{\mathbf{y}^T\mathbf{y}}{2}\right] \mathrm{d}\mathbf{y} = \epsilon f''_{ij}(\mathbf{0})\Xi_{ij},
\end{aligned}$$

the Frobenius product of the Hessian matrix and covariant matrix $\mathbf{\Xi}$,

$$\begin{aligned}
&\left[\left(2\pi\epsilon\right)^n \det\left(\mathbf{\Xi}\right)\right]^{-\frac{1}{2}} \int_{\mathbb{R}^n} f'''_{ijk}(\mathbf{0})x_i x_j x_k \exp\left[-\frac{1}{2\epsilon}\mathbf{x}^T\mathbf{\Xi}\mathbf{x}\right] \mathrm{d}\mathbf{x} = 0, \\
&\left[\left(2\pi\epsilon\right)^n \det\left(\mathbf{\Xi}\right)\right]^{-\frac{1}{2}} \int_{\mathbb{R}^n} f''''_{ijk\ell}(\mathbf{0})x_i x_j x_k x_\ell \exp\left[-\frac{1}{2\epsilon}\mathbf{x}^T\mathbf{\Xi}\mathbf{x}\right] \mathrm{d}\mathbf{x} \\
= \quad & \epsilon^2 f''''_{ijk\ell}(\mathbf{0})\Xi^{\frac{1}{2}}_{i\mu}\Xi^{\frac{1}{2}}_{j\nu}\Xi^{\frac{1}{2}}_{k\rho}\Xi^{\frac{1}{2}}_{\ell\kappa}\Theta_{\mu\nu\rho\kappa}, \\
&\left[\left(2\pi\epsilon\right)^n \det\left(\mathbf{\Sigma}\right)\right]^{-\frac{1}{2}} \int_{\mathbb{R}^n} f'''_{ijk}(\mathbf{0})x_i^2 x_j^2 x_k^2 \exp\left[-\frac{1}{2\epsilon}\mathbf{x}^T\mathbf{\Xi}\mathbf{x}\right] \mathrm{d}\mathbf{x} \\
= \quad & \epsilon^3\left(2\pi\right)^{-\frac{n}{2}} f'''_{ijk}(\mathbf{0})\Xi^{-\frac{1}{2}}_{i\mu}\Xi^{-\frac{1}{2}}_{i\mu'}\Xi^{-\frac{1}{2}}_{j\nu}\Xi^{-\frac{1}{2}}_{j\nu'}\Xi^{-\frac{1}{2}}_{k\rho}\Xi^{-\frac{1}{2}}_{k\rho'} \int_{\mathbb{R}^n} y_\mu y_{\mu'} y_\nu y_{\nu'} y_\rho y_{\rho'} \exp\left[-\frac{\mathbf{y}^T\mathbf{y}}{2}\right] \mathrm{d}\mathbf{x} \\
= \quad & \epsilon^3 f'''_{ijk}(\mathbf{0})\Xi^{-\frac{1}{2}}_{i\mu}\Xi^{-\frac{1}{2}}_{i\mu'}\Xi^{-\frac{1}{2}}_{j\nu}\Xi^{-\frac{1}{2}}_{j\nu'}\Xi^{-\frac{1}{2}}_{k\rho}\Xi^{-\frac{1}{2}}_{k\rho'}\Lambda_{\mu\mu'\nu\nu'\rho\rho'}.
\end{aligned}$$

Applying (6.27) to both numerator and denominator of the lhs of (6.28),

$$\frac{f(\mathbf{x}^*) + \epsilon\left\{\begin{array}{c} \frac{f''_{ij}(\mathbf{x}^*)\Xi_{ij}}{2} - \left[\frac{f'_i(\mathbf{x}^*)h'''_{jk\ell}(\mathbf{x}^*)}{6} + \frac{f_\alpha(\mathbf{x}^*)h''''_{ijk\ell}(\mathbf{x}^*)}{24}\right]\Xi^{\frac{1}{2}}_{i\mu}\Xi^{\frac{1}{2}}_{j\nu}\Xi^{\frac{1}{2}}_{k\rho}\Xi^{\frac{1}{2}}_{\ell\kappa}\Theta_{\mu\nu\rho\kappa} \\ +\frac{f(\mathbf{x}^*)[h'''_{ijk}(\mathbf{x}^*)]^2}{72}\Xi^{-\frac{1}{2}}_{i\mu}\Xi^{-\frac{1}{2}}_{i\mu'}\Xi^{-\frac{1}{2}}_{j\nu}\Xi^{-\frac{1}{2}}_{j\nu'}\Xi^{-\frac{1}{2}}_{k\rho}\Xi^{-\frac{1}{2}}_{k\rho'}\Lambda_{\mu\mu'\nu\nu'\rho\rho'} \end{array}\right\} + O(\epsilon^2)}{1 + \epsilon\left\{\begin{array}{c} -\left[\frac{h''''_{ijk\ell}(\mathbf{x}^*)}{24}\right]\Xi^{\frac{1}{2}}_{i\mu}\Xi^{\frac{1}{2}}_{j\nu}\Xi^{\frac{1}{2}}_{k\rho}\Xi^{\frac{1}{2}}_{\ell\kappa}\Theta_{\mu\nu\rho\kappa} \\ +\frac{[h'''_{ijk}(\mathbf{x}^*)]^2}{72}\Xi^{-\frac{1}{2}}_{i\mu}\Xi^{-\frac{1}{2}}_{i\mu'}\Xi^{-\frac{1}{2}}_{j\nu}\Xi^{-\frac{1}{2}}_{j\nu'}\Xi^{-\frac{1}{2}}_{k\rho}\Xi^{-\frac{1}{2}}_{k\rho'}\Lambda_{\mu\mu'\nu\nu'\rho\rho'} \end{array}\right\} + O(\epsilon^2)}$$

$$= f(\mathbf{x}^*) + \epsilon\left[\frac{f''_{ij}(\mathbf{x}^*)\Xi_{ij}}{2} - \frac{f'_i(\mathbf{x}^*)h'''_{jk\ell}(\mathbf{x}^*)\Xi^{\frac{1}{2}}_{i\mu}\Xi^{\frac{1}{2}}_{j\nu}\Xi^{\frac{1}{2}}_{k\rho}\Xi^{\frac{1}{2}}_{\ell\kappa}\Theta_{\mu\nu\rho\kappa}}{6}\right] + O(\epsilon^2).$$

□

Proof of Lemma 6.2.3

We only provide the proof for the case $\mathbf{x} \in \mathbb{R}^1$, which is denoted by x. For higher dimensions, results are the same by using the notations from Lemma 6.2.2.

Proof. Let the global minimum of $[h(x) - \epsilon \ln g(x)]$ be at $\tilde{x}^* = x^* + \Delta x(\epsilon)$. Clearly, $\Delta x \to 0$ as $\epsilon \to 0$. In fact,

$$\begin{aligned}
&\left[h'(x) - \epsilon\left(\frac{g'(x)}{g(x)}\right)\right]_{x=x^*+\Delta x} = 0, \\
&h'(x^*) + h''(x^*)\Delta x - \epsilon\left(\frac{g'(x^*)}{g(x^*)}\right) = 0, \\
&\Delta x = \epsilon\left(\frac{g'(x^*)}{g(x^*)h''(x^*)}\right) + O\big(\epsilon^2\big).
\end{aligned}$$

Now we apply the Eq. (6.28) in Lemma 6.2.2:

$$\begin{aligned}
&\frac{\displaystyle\int_{-\infty}^{\infty} f(x)g(x)e^{-\frac{h(x)}{\epsilon}}\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty} g(x)e^{-\frac{h(x)}{\epsilon}}\mathrm{d}x} = \frac{\displaystyle\int_{-\infty}^{\infty} f(x)e^{-\frac{h(x)-\epsilon\ln g(x)}{\epsilon}}\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty} e^{-\frac{h(x)-\epsilon\ln g(x)}{\epsilon}}\mathrm{d}x} \\
= \quad & f(\tilde{x}^*) + \epsilon\left[\frac{f''(\tilde{x}^*)}{2h''(\tilde{x}^*)} - \frac{f'(\tilde{x}^*)h'''(\tilde{x}^*)}{2[h''(\tilde{x}^*)]^2}\right] + O(\epsilon^2), \\
= \quad & f(x^*) + \epsilon\left(\underbrace{\frac{f'(x^*)g'(x^*)}{g(x^*)h''(x^*)}}_{f'(x^*)\Delta x \text{ due to } g(x)} + \underbrace{\frac{f''(x^*)}{2h''(x^*)}}_{\text{due to } f(x)} - \underbrace{\frac{f'(x^*)h'''(x^*)}{2[h''(x^*)]^2}}_{\text{due to non-quadratic } h(x)}\right) + O(\epsilon^2).
\end{aligned} \tag{6.119}$$

For the terms on the order of ϵ, replacing $\tilde{x}^*$ by x^* only affects the order ϵ^2 term.

□

Proof of Lemma 6.6.1

Here we provide an additional Lemma, which is not used in the present work, but it is useful for related fields.

Lemma 6.6.1.

$$\frac{\displaystyle\int_{-\infty}^{\infty} f^2(x)g(x)e^{-\frac{h(x)}{\epsilon}}\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty} g(x)e^{-\frac{h(x)}{\epsilon}}\mathrm{d}x} - \left[\frac{\displaystyle\int_{-\infty}^{\infty} f(x)g(x)e^{-\frac{h(x)}{\epsilon}}\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty} g(x)e^{-\frac{h(x)}{\epsilon}}\mathrm{d}x}\right]^2 = \epsilon\left(\frac{f'^2(x^*)}{h''(x^*)}\right) + O\big(\epsilon^2\big). \tag{6.120}$$

Proof.

$$\frac{\displaystyle\int_{-\infty}^{\infty} f^2(x)g(x)e^{-\frac{h(x)}{\epsilon}}\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty} g(x)e^{-\frac{h(x)}{\epsilon}}\mathrm{d}x} - \left[\frac{\displaystyle\int_{-\infty}^{\infty} f(x)g(x)e^{-\frac{h(x)}{\epsilon}}\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty} g(x)e^{-\frac{h(x)}{\epsilon}}\mathrm{d}x}\right]^2$$
$$\begin{aligned}
&= \epsilon\left(\frac{2f(x^*)f'(x^*)g'(x^*)}{g(x^*)h''(x^*)} + \frac{2f'^2(x^*) + 2f(x^*)f''(x^*)}{2h''(x^*)} - \frac{2f(x^*)f'(x^*)h'''(x^*)}{2[h''(x^*)]^2}\right)\\
&- 2\epsilon f(x^*)\left(\frac{f'(x^*)g'(x^*)}{g(x^*)h''(x^*)} + \frac{f''(x^*)}{2h''(x^*)} - \frac{f'(x^*)h'''(x^*)}{2[h''(x^*)]^2}\right) + O(\epsilon^2)\\
&= \epsilon\left(\frac{f'^2(x^*)}{h''(x^*)}\right) + O(\epsilon^2).
\end{aligned} \tag{6.121}$$

□

6.6.2 Proof of Theorem 6.3.4

Proof. Given $\mathbf{x_1} \in \Gamma$, let $\mathcal{P}_1$ be a plane containing $\mathbf{x_1}$ and perpendicular to the vector $\boldsymbol{\gamma}(\mathbf{x_1})$. Given another point $\mathbf{x}_2 \in \Gamma$, let $\mathcal{P}_2$ be a plane perpendicular to the vector $\boldsymbol{\gamma}(\mathbf{x}_2)$. Let $\mathcal{S}_1 \subset \mathcal{P}_1$ and $\mathcal{S}_2 \subset \mathcal{P}_2$ be compact sets such that

$$\max_{\mathbf{x},\mathbf{y}\in\mathcal{S}_1} ||\mathbf{x}-\mathbf{y}|| = \max_{\mathbf{x},\mathbf{y}\in\mathcal{S}_2} ||\mathbf{x}-\mathbf{y}|| = \delta > 0. \tag{6.122}$$

We then can define a tube $\Phi(\delta)$ with two side boundaries $\mathcal{S}_1$ and $\mathcal{S}_2$ and $\Gamma \subset \Phi(\delta)$.

Recall that $\boldsymbol{\gamma}_\epsilon(\mathbf{x}) = \pi_\epsilon(\mathbf{x})^{-1}\mathbf{J}[\pi_\epsilon(\mathbf{x})]$. By the stationary Fokker-Planck equation, we have that

$$\nabla\cdot(\boldsymbol{\gamma}_\epsilon(\mathbf{x})\pi_\epsilon(\mathbf{x})) = 0, \quad \text{for all } \mathbf{x}\in\mathbb{R}^n. \tag{6.123}$$

Furthermore, by the Gauss's theorem,

$$\int_{\mathcal{S}} \big(\boldsymbol{\gamma}_\epsilon(\mathbf{x})\pi_\epsilon(\mathbf{x})\cdot\mathbf{n}\big)\mathrm{d}\mathcal{S} = \int_{\mathcal{V}} \big(\nabla\cdot(\boldsymbol{\gamma}_\epsilon(\mathbf{x})\pi_\epsilon(\mathbf{x})\big)\mathrm{d}\mathcal{V} = 0, \tag{6.124}$$

where $\mathcal{S}$ in the surface of $\Phi(\delta)$, $\mathbf{n}$ is the outward normal vector to $\mathcal{S}$, and $\mathcal{V}$ is the volume of $\Phi(\delta)$.

For the left hand side of Eq. (6.124), it can be written as a sum of three terms

$$\begin{aligned}
\int_{\mathcal{S}} \big(\boldsymbol{\gamma}_\epsilon(\mathbf{x})\pi_\epsilon(\mathbf{x})\cdot\mathbf{n}\big)\mathrm{d}\mathcal{S} = &\int_{\mathcal{S}_1} \big(\boldsymbol{\gamma}_\epsilon(\mathbf{x})\pi_\epsilon(\mathbf{x})\cdot\mathbf{n}\big)\mathrm{d}\mathcal{S}_1 + \int_{\mathcal{S}_2} \big(\boldsymbol{\gamma}_\epsilon(\mathbf{x})\pi_\epsilon(\mathbf{x})\cdot\mathbf{n}\big)\mathrm{d}\mathcal{S}_2\\
&+ \int_{\mathcal{S}_3} \big(\boldsymbol{\gamma}_\epsilon(\mathbf{x})\pi_\epsilon(\mathbf{x})\cdot\mathbf{n}\big)\mathrm{d}\mathcal{S}_3,
\end{aligned} \tag{6.125}$$

in which $\mathcal{S}_3$ is the lateral surface of $\Phi(\delta)$ and $\mathcal{S}_3 \cap \Gamma = \emptyset$.

For every $\mathbf{y} \in \Gamma$, we denote $\mathcal{S}_\mathbf{y} := \Phi(\delta) \cap \mathcal{P}_\mathbf{y}$, $\mathcal{P}_\mathbf{y}$ is the plane perpendicular to $\boldsymbol{\gamma}(\mathbf{y})$. By Lemma 6.2.2,

$$\frac{\int_{\mathcal{S}_\mathbf{y}} f(\mathbf{x}) e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathcal{S}_\mathbf{y}}{\int_{\mathcal{S}_\mathbf{y}} e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathcal{S}_\mathbf{y}} = f(\mathbf{y}) + O(\epsilon), \tag{6.126}$$

for any continuous and bounded function $f : \mathbb{R}^n \to \mathbb{R}$. Furthermore, by the definition of function v in Lemma 6.3.4, we can approximate the ratio of two integrals

$$\frac{\int_{\mathcal{S}_\mathbf{y}} e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathcal{S}_\mathbf{y}}{\int_{\mathcal{S}_1} e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathcal{S}_1} = \frac{\sqrt{2\pi\epsilon v(\mathbf{y})} e^{\frac{-\varphi(\mathbf{y})}{\epsilon}}}{\sqrt{2\pi\epsilon v(\mathbf{x})} e^{\frac{-\varphi(\mathbf{x}_1)}{\epsilon}}} + O(\epsilon) = \frac{\sqrt{v(\mathbf{y})}}{\sqrt{v(\mathbf{x}_1)}} + O(\epsilon), \tag{6.127}$$

in which we use Laplace's method in the first equality and $\varphi \equiv 0$ on Γ in the second equality. To choose $f(\mathbf{x}) = \omega(\mathbf{x})$ for Eq. (6.126), combined with the result of (6.127), we can obtain

$$\frac{\int_{\mathcal{S}_\mathbf{y}} \omega(\mathbf{x}) e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathcal{S}_\mathbf{y}}{\int_{\mathcal{S}_1} e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathcal{S}_1} = \frac{\omega(\mathbf{y})\sqrt{v(\mathbf{y})}}{\sqrt{v(\mathbf{x}_1)}} + O(\epsilon). \tag{6.128}$$

Since the function $e^{-\frac{\varphi(\mathbf{x})}{\epsilon}}$ is concentrated near Γ, we can further approximate the normalization factor $\int_{\mathbb{R}^n} \omega(\mathbf{x}) e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathbf{x}$ as follows

$$\frac{\int_{\mathbb{R}^n} \omega(\mathbf{x}) e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathbf{x}}{\int_{\mathcal{S}_1} e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathcal{S}_1} = \frac{\int_\Gamma \int_{\mathcal{S}_\mathbf{y}} \omega(\mathbf{x}) e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathcal{S}_\mathbf{y} \mathrm{d}\mathbf{y}}{\int_{\mathcal{S}_1} e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathcal{S}_1} + O(\epsilon). \tag{6.129}$$

By (6.128) and (6.129), we have that

$$\frac{\int_{\mathbb{R}^n} \omega(\mathbf{x}) e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathbf{x}}{\int_{\mathcal{S}_1} e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathcal{S}_1} = \frac{\int_\Gamma \omega(\mathbf{y})\sqrt{v(\mathbf{y})} \mathrm{d}\mathbf{y}}{\sqrt{v(\mathbf{x}_1)}} + O(\epsilon). \tag{6.130}$$

By the WKB expansion of π_ϵ in Eq. (6.55), with Eq. (6.130), the first term on the right hand side of Eq. (6.125) can be written as

$$\begin{aligned} \int_{\mathcal{S}_1} \big(\boldsymbol{\gamma}_\epsilon(\mathbf{x})\pi_\epsilon(\mathbf{x}) \cdot \mathbf{n}\big) \mathrm{d}\mathcal{S}_1 &= \frac{\int_{\mathcal{S}_1} (\boldsymbol{\gamma}_\epsilon(\mathbf{x}) \cdot \mathbf{n})\, \omega(\mathbf{x}) e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathcal{S}_1}{\int_{\mathbb{R}^n} \omega(\mathbf{x}) e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathbf{x}} \\ &= \left(\frac{\sqrt{v(\mathbf{x}_1)}}{\int_\Gamma \omega(\mathbf{y})\sqrt{v(\mathbf{y})} \mathrm{d}\mathbf{y}} \right) \left(\frac{\int_{\mathcal{S}_1} (\boldsymbol{\gamma}_\epsilon(\mathbf{x}) \cdot \mathbf{n})\, \omega(\mathbf{x}) e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathcal{S}_1}{\int_{\mathcal{S}_1} e^{\frac{-\varphi(\mathbf{x})}{\epsilon}} \mathrm{d}\mathcal{S}_1} \right) + O(\epsilon). \end{aligned} \tag{6.131}$$

Note that $\boldsymbol{\gamma}_\epsilon(\mathbf{x}) \to \boldsymbol{\gamma}(\mathbf{x})$. Without loss of generality, we assume that $\boldsymbol{\gamma}(\mathbf{x}_1)$ is inflow and $\boldsymbol{\gamma}(\mathbf{x}_2)$ is outflow of $\Phi(\delta)$. To choose $f(\mathbf{x}) = (\boldsymbol{\gamma}_\epsilon(\mathbf{x}) \cdot \mathbf{n})\, \omega(\mathbf{x})$ for Eq. (6.126), combined with Eq. (6.131), we then obtain

$$\int_{\mathcal{S}_1} \big(\boldsymbol{\gamma}_\epsilon(\mathbf{x})\pi_\epsilon(\mathbf{x}) \cdot \mathbf{n}\big) \mathrm{d}\mathcal{S}_1 \to -C\sqrt{v(\mathbf{x}_1)}\omega(\mathbf{x}_1)||\boldsymbol{\gamma}(\mathbf{x}_1)|| \quad \text{as} \quad \epsilon \to 0, \tag{6.132}$$

in which the constant $C = 1/\int_\Gamma \omega(\mathbf{y})\sqrt{v(\mathbf{y})}\mathrm{d}\mathbf{y}$. By the same approach, the second term on the right hand side of Eq. (6.125) has a convergence

$$\int_{\mathcal{S}_2} \big(\boldsymbol{\gamma}_\epsilon(\mathbf{x})\pi_\epsilon(\mathbf{x}) \cdot \mathbf{n}\big) \mathrm{d}\mathcal{S}_2 \to C\sqrt{v(\mathbf{x}_2)}\omega(\mathbf{x}_2)||\boldsymbol{\gamma}(\mathbf{x}_2)|| \quad \text{as} \quad \epsilon \to 0. \tag{6.133}$$

Since $\mathcal{S}_3 \cap \Gamma = \emptyset$, the third term

$$\int_{\mathcal{S}_3} \big(\boldsymbol{\gamma}_\epsilon(\mathbf{x})\pi_\epsilon(\mathbf{x}) \cdot \mathbf{n}\big) \mathrm{d}\mathcal{S}_3 \to 0 \quad \text{as} \quad \epsilon \to 0. \tag{6.134}$$

To apply the results (6.132), (6.133), and (6.134) to the equations (6.124) and (6.125), we can show that

$$\Big| C\sqrt{v(\mathbf{x}_1)}\omega(\mathbf{x}_1)||\boldsymbol{\gamma}(\mathbf{x}_1)|| - C\sqrt{v(\mathbf{x}_2)}\omega(\mathbf{x}_2)||\boldsymbol{\gamma}(\mathbf{x}_2)|| \Big| = 0. \tag{6.135}$$

Since Eq. (6.135) holds for every pair of two points on Γ, $\sqrt{v(\mathbf{x})}\omega(\mathbf{x})||\boldsymbol{\gamma}(\mathbf{x})||$ is constant on Γ.

For $\mathbf{x} \in \Gamma$, the marginal density can be approximated by

$$g_\epsilon(\mathbf{x}) = \frac{\int_{\mathbb{R}^n \setminus \Gamma} \omega(\mathbf{y}) e^{\frac{-\varphi(\mathbf{y})}{\epsilon}} \mathrm{d}\mathbf{y}}{\int_{\mathbb{R}^n} \omega(\mathbf{y}) e^{\frac{-\varphi(\mathbf{y})}{\epsilon}} \mathrm{d}\mathbf{y}} = \frac{\int_{\mathcal{S}_\mathbf{x}} \omega(\mathbf{y}) e^{\frac{-\varphi(\mathbf{y})}{\epsilon}} \mathrm{d}\mathcal{S}_\mathbf{x}}{\int_{\mathbb{R}^n} \omega(\mathbf{y}) e^{\frac{-\varphi(\mathbf{y})}{\epsilon}} \mathrm{d}\mathbf{y}} + O(\epsilon) = \frac{\omega(\mathbf{x})\sqrt{v(\mathbf{x})}}{\int_\Gamma \omega(\mathbf{y})\sqrt{v(\mathbf{y})}\mathrm{d}\mathbf{y}} + O(\epsilon), \tag{6.136}$$

which follows the steps from Eq. (6.126) to Eq. (6.131) with choosing $\boldsymbol{\gamma}_\epsilon(\mathbf{x}) \cdot \mathbf{n} = 1$. Furthermore, since $\sqrt{v(\mathbf{x})}\omega(\mathbf{x})||\gamma(\mathbf{x})||$ is constant on Γ, with the result (6.136), there exists a constant K such that

$$g_\epsilon(\mathbf{x})||\boldsymbol{\gamma}(\mathbf{x})|| = K + O(\epsilon). \tag{6.137}$$

□

Chapter 7

CONCLUSIONS AND FUTURE WORK

In part I, a new limit theorem in conditional probability distribution is introduced to justify J. W. Gibbs' statistical canonical ensemble theory for heterogeneous systems and to further generalize the Boltzmann distribution law for strongly coupled systems. As for its applications to biophysical chemistry [97], consider a single RNA molecule immersed in one mole of aqueous solution and the RNA molecule has interactions with molecules in the solution. In this case, the RNA itself is a macromolecule and the aqueous solution has molecules in the order 10^{23}. With the extremely large degrees of freedom, obviously, it is a complex system; in contrast, the distribution of the state of RNA in this enormously complicated system simply follow the Boltzmann distribution law.

In classical statistical mechanics, nearly non-interacted homogeneous systems, e.g. ideal gas systems, also follow the Boltzmann distribution law [92]. However, I shall emphasize that the Boltzmann distribution law for strongly coupled heterogeneous systems is a "modified" Boltzmann distribution law with the pseudo-Hamiltonian and the modified temperature including the correlation between the RNA and the solution. By the theory of ensemble equivalence, the existence of pseudo-Hamiltonian is guaranteed by the existence of Hamiltonian of mean force; however, once the equivalence of ensemble is invalid when the system has a critical phenomenon, then the existence of the pseudo-Hamiltonian needs a rigorous proof in future work. Furthermore, the modified temperate with non-negligible correlations might be corresponding to the long-range interactions between the RNA molecule to each molecule of the solution. In the present work, it is only a hypothesis and further mechanistic investigations are required.

Here is another example for the application of the canonical ensemble theory, again from biophysical chemistry: the Boltzmann distribution of side-chain conformation in proteins has been widely used and checked with experimental data [24, 212]. In future work, by comparing the modified Boltzmann with those data, we can further verify whether or not the modified Boltzmann can precisely predict the distribution of this example when the side chain has strong interactions with the solvent. More importantly, the mathematical theory provided in chapter 2 is a theory of conditional probability, so it can be considered as a powerful tool to analyze large-scale data of side-chain conformation in proteins [136]. In particular, the side-chain conformation change has be considered as a result of protein mutations [132], so this system turns out to be a non-classical canonical ensemble system, in which is lack of a well-defined Hamiltonian of the protein.

In this case, the emergent law of conditional probability still can justify that Boltzmann law for the side-chain conformation distribution from the perspective of data science by three steps: First, we need to carefully examine the "constraint" of those data corresponding to the "condition" of the total system. Second, we should check that the protein has to be relatively small to the solvent in terms of certain quantity, which turns out to be the "effective" Hamiltonian of the system. Third, if the protein and the solvent have strong correlations, then there will be an additional term in the temperature related to the mutual information of the protein and the solvent. Beyond biophysical chemistry, following those three steps, the generalized canonical ensemble theory from the perspective of date science can also be applied to the following two fields [32]: (i) Economics: the distribution law of cash possessed by a city of a rather large economic system; and (ii) Ecology: the distribution law of a population contained in a much large ecological system. In both of the examples, the total system can be assumed as a constant with respect to a proper time scale.

In part II, a landscape theory of deterministic dynamical systems is introduced. That landscape theory is a "recipe" for how to find a Lyapunov function for a dynamical system even the flow of this system does not follow the gradient of a potential function. Interestingly, this recipe to find that Lyapunov function to characterize the "deterministic" dynamics is by random perturbations with small "noises" parametrized by ϵ . As the noises disappear ($\epsilon \to 0$) and the system reaches its steady state ($t \to 0$), an ϵ, t independent function named the large-deviation rate function emerges and serves as the Lyapunov function for the deterministic dynamics. Using a dynamical system to model a biological phenomenon prevails among applied mathematical point of view to characterize the evolution of the system (transient dynamics) and even further predict the destiny of the system (stead state analysis). Once a landscape is constructed for a biological system, we are able to understand the biological phenomenon from a higher perspective, which is similar to the concept of energy, entropy, and free energy in physics. Therefore, this is a perfect example following the essence of this dissertation: a simple law as a consequence of certain limits in mathematics can be applied to justify an universal law in natural science.

In addition, since noise is ubiquitous in complex living systems, perturbing a deterministic dynamical system with small noises is not only a mathematical technique to find the Lyapunov function but also a more realistic description of the biological phenomenon. The scaling hypothesis introduced in chapter 6 serves as a scientific theory: parameter ϵ for controlling the size of noises is a corollary of the observer's scopes in both space and time. As an "imperfect" measurement adopted by the observer is too coarse to measure the ϵ size of noises, the observed dynamics becomes "perfectly" deterministic and it is precisely characterized the emergent large-deviation rate function. This idea echos Anderson's statement: "solve a problem with infinitely many bodies, and then apply the result to a finite system", which is introduced in chapter 1.

The landscape of a system with its steady state as a stable limit cycle requires careful analysis since its Lyapunov function is only in a weak sense: the dynamics on the function is no longer strictly monotonic. There is a "flat" ring on the bottom of the Lyapunov function and therefore we are lost on the cycle for lack of knowledge about the direction and the speed of flow. This is the reason that the prefactor which characterizes the landscape in the next order plays an important role on the limit cycle. It portrays the dynamics on small local hills on top of the Lyapunov function, and the heights of those hills are intimately related to the speed on the cycle. In contradistinction to the well-established theories of the Hamilton-Jacobi equation for the large deviation rate function, complete mathematical analysis of perfactor in the WKB method is missing and it is worthy of attention in the future. In applications, this landscape theory including both the leading order and the second order terms for stable limit cycles can help us to paint a clearer picture of oscillatory behaviors in biology, e.g. (i) cell-division cycle in cell biology [124], (ii) spike patterns in neuroscience [59], and (iii) predator-prey interactions in ecology [135].

Following from the theories in this dissertation for complex systems with intrinsic thermodynamic phenomena in the mesoscopic point of view, I want to introduce two pillars of life: complexity and open systems. The former is associated with uncertainty, chaos, and many-body interactions and is represented by the theory of probability; The later is related to the notion of nonequilibrium steady state in stochastic dynamics. The probabilistic interpretation of nonequilibrium thermodynamics from a mesoscopic point of view lays the basis for a mathematical description of life. In certain scaling limits, the emergent behaviors illustrate the existence of simple macroscopic laws. Based on those two pillars, next sections of future work are about my hypotheses for complex living systems: (i) A stochastic coupled diffusion model for circulatory system quantitatively accounts for the energetics and entropies in the dynamics of blood flow. (ii) Sufficient free energy supply from ATP hydrolysis plays a pivotal role in negative feedback in physiology for homeostasis.

7.1 A stochastic analysis of a coupled diffusion model for cardiovascular system

A cardiovascular system can be simply modeled by three components and three connections. The three components: the heart, the artery, and the vein - all can be treated as elastic tubes. Each elastic tube is characterized by *linear elasticity* with its natural volume. Active contraction of cardiac muscle is represented by the ability of the heart to change its elasticity, which is an emergent phenomenon - a collective behaviors of the release of calcium from cardiac muscles triggered by propagation of action potential for a short period of time [1]. High elasticity represents the *phase of contraction* of the heart; Low elasticity represents the *phase of relaxation* of the heart. Three connections are the inflow valve, the outflow valve and microvessels, respectively. The inflow valve only allows one-way flow from the vein to the heart; The outflow valve only allows one-way flow from the heart to the artery; Microvessles allow two-way flow between the artery and the vein with equal resistances. By this toy model, we are able to describe continuous blood flows generated by alternations of the systolic phase (heart contraction) and the diastolic phase (heart relaxation). The whole process is illustrated by Figures 7.1 and 7.2. In order to build a solid concept of *macroscopic entropy production* in the process, we apply stochastic analysis of a *couple*

[1]Even though there is time lag of the propagation between different parts of the heart, it is negligible with respect to the measurable time interval.

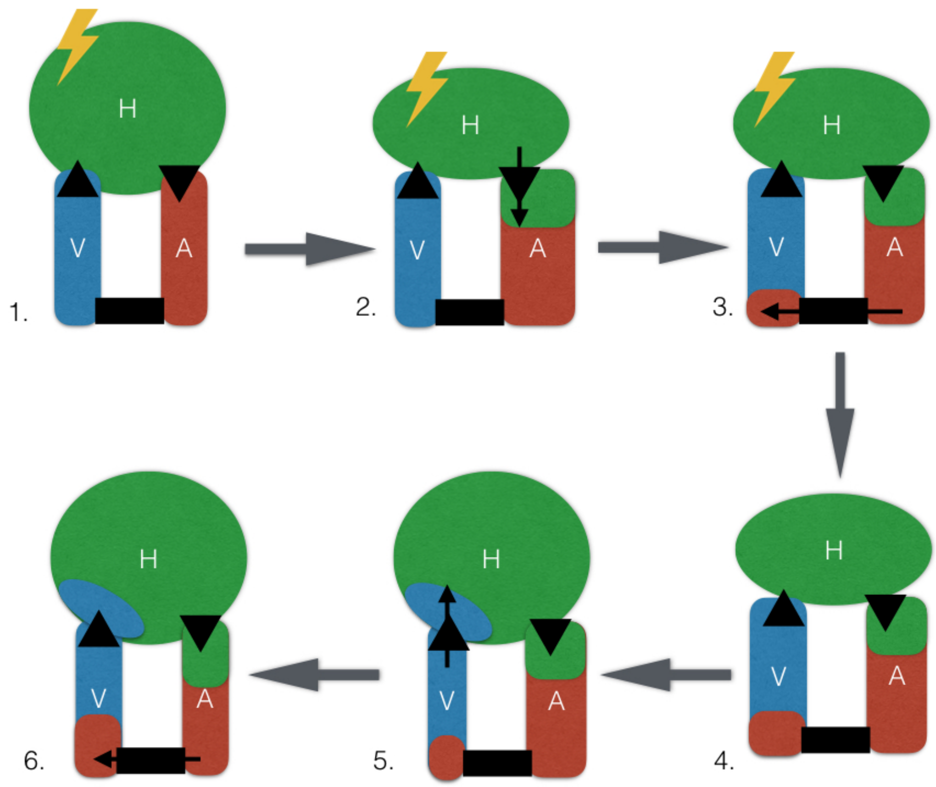

Figure 7.1: The process starts from an equilibrium state. 1. The heart is in a diastolic phase (low elasticity) and the three components have equal blood pressures with no blood flow through each connection. 2. At some point, heart contraction begins. 3. Once blood flows into the artery, there will be a blood pressure difference between the artery and the vein, then it generates blood flows through the microvessels. 4. The system is reaching to its new equilibrium of the systolic phase (high elasticity). 5. The heart starts relaxing so its elasticity decreases to its original low value immediately. 6. Once blood flows into the heart from the vein, there is a result of a blood pressure difference between the artery and the vein, then it generates blood flows through the microvessels. After completing one entire process, the three elastic tubes and connections return back to their original states. However, if we label the blood by a color, we are able to observe blood flow in a clockwise direction.

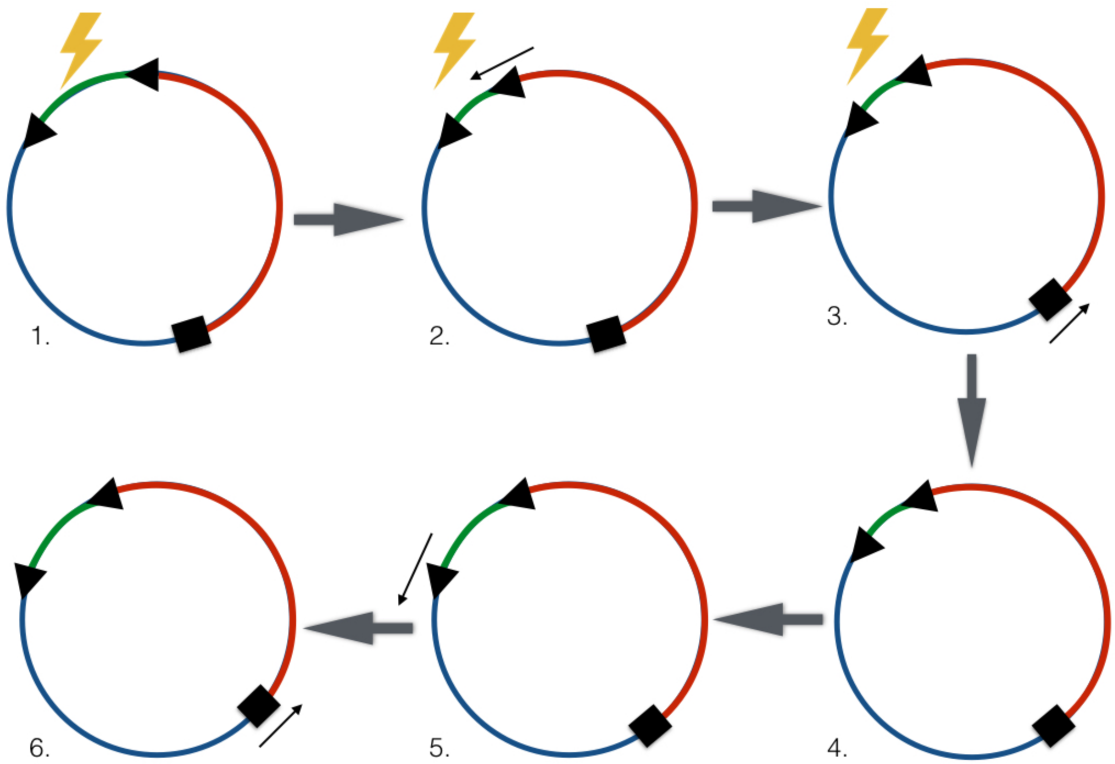

Figure 7.2: A Map from a static coordinate in Figure 7.1 to a moving coordinate: The three elastic tubes are represented by three springs. The volumes of tubes are mapped to the lengths of springs, respectively. Since the total volume of blood is conserved in our model, the entire length of three springs is fixed on a circle. The three connections are mapped to three overdamped particles. Blood flowing through a resistant valve and a corresponding particle flowing in the viscous blood are two sides of the same coin, which just depends on the observer. Therefore, the friction for the mapped overdamped particle is corresponding to the resistance to the flow of blood across the valve. For an one-way valve, the friction for its mapped particle is very high while moving clockwise in the blood. On the other hand, it is very low while moving counter-clockwise. For microcirculation, since it is two-way, the friction of its mapped particle is equal while moving in each direction.

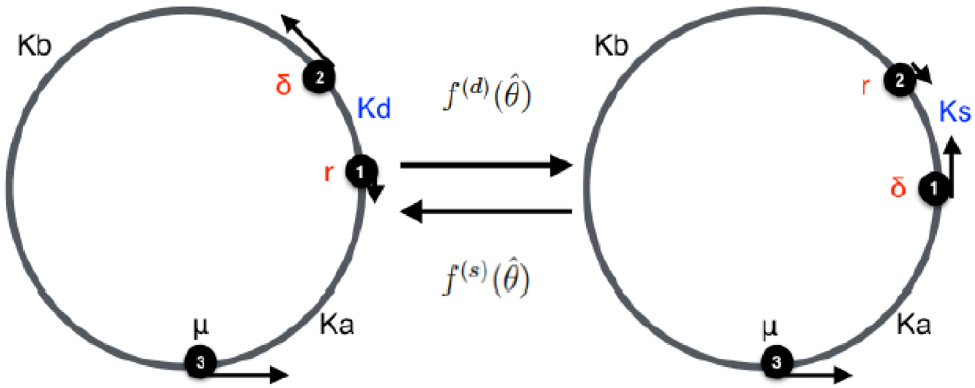

Figure 7.3: The system is composed by three springs with natural lengths l, l_a, l_b and elasiticities $k_d(k_s), k_a, k_b$ who are connected by three overdamped particles with friction r, δ, μ. One of the springs representing the heart in the model has alternative elasticities $k_d \leq k_s$, in which k_d is corresponding to the diastolic phase and k_s is corresponding to the heart systolic phase.

diffusion model [155] to our toy model of the cardiovascular system with five steps:

Step 1: We introduce a *switched deterministic dynamical system* for the human cardiovascular system illustrated by Figure 7.3. Assume the system starts from a diastolic phase and three springs are in an equilibrium state. At a certain moment, the heart is activated by an electrical impulse, its elasticity jumps from k_d to k_s immediately, which means the system enters into the systolic phase. The switched dynamical system is as following:

$$
\begin{aligned}
r\frac{d\theta_1}{dt} &= k_d(\theta_2 - \theta_1 - l) - k_a((2\pi - (\theta_3 - \theta_1) - l_a), \\
\delta\frac{d\theta_2}{dt} &= k_b(\theta_3 - \theta_2 - l_b) - k_s(\theta_2 - \theta_1 - l), \\
\mu\frac{d\theta_3}{dt} &= k_a((2\pi - (\theta_3 - \theta_1) - l_a) - k_b(\theta_3 - \theta_2 - l_b),
\end{aligned}
\tag{7.1}
$$

$$
\begin{aligned}
\delta\frac{d\theta_1}{dt} &= k_s(\theta_2 - \theta_1 - l) - k_a((2\pi - (\theta_3 - \theta_1) - l_a), \\
r\frac{d\theta_2}{dt} &= k_b(\theta_3 - \theta_2 - l_b) - k_d(\theta_2 - \theta_1 - l), \\
\mu\frac{d\theta_3}{dt} &= k_a((2\pi - (\theta_3 - \theta_1) - l_a) - k_b(\theta_3 - \theta_2 - l_b).
\end{aligned}
\tag{7.2}
$$

As our narrative above, an initial condition is given to Eq. (7.1) (diastolic phase) at the beginning, then the system switches to Eq. (7.2) (systolic phase) when depolarization of the cardiac muscles occurs at the sinus node.

Step 2: For the switched dynamical system, in order to guarantee Eq. (7.1) is well-defined in the *diastolic phase* and Eq. (7.2) is well-defined in the *systolic phase*, respectively (i.e. the set of parameters should guarantee the correct direction of flows), we need to find sufficient conditions such that the dynamics of $\theta_1, \theta_2, \theta_3$ are monotonic during each phase $[t_i, t_{i+1})$ regardless of the pattern of alternating periods.

Theorem 7.1.1. *(Uni-directional flows) If the parameters in dynamics (7.1, 7.2) satisfy*

$$r > \max(\frac{k_d}{k_b}\mu,\ \frac{k_a}{k_b}\delta, \frac{k_s}{k_a}\mu,\ \frac{k_b}{k_a}\delta), \tag{7.3}$$

then the signs of $\frac{d\theta_1}{dt}(t)$, $\frac{d\theta_2}{dt}(t)$, $\frac{d\theta_3}{dt}(t)$ *are invariant for* $[t_i, t_{i+1})$*, where* $t_i \geq 0$ *is every moment of perturbations (change of parameters) for the system. Furthermore,* $\frac{d\theta_3}{dt}(t) \geq 0$ *for all* $t \geq 0$.

Theorem 7.1.1 shows that θ_3 moves uni-directionally no matter what the initial condition is or when the alternation occurs. In normal human physiology [105], the condition (7.3) is usually satisfied: it guarantees that unidirectional systemic blood flow can not be violated for any pattern of heatbeat intervals. See our simulation for $\theta_3(t)$ in Figure 7.4 and $d\theta_3/dt$ in Figure 7.5. Not surprisingly, our simulation with generally approved parameters [105] is consistent with well-accepted experimental results of hemodynamics in human physiology [211].

Step 3: As we discussed previously, deterministic laws are observed in macroscopy. In order to rigorously define *entropy production rate* for the human cardiovascular system, we need to analyze this system on *the mesoscopic scale* with uncertainty. It says we add noise to Eqs. (7.1) and (7.2), to obtain *a switched stochastic dynamical system*. In this process, randomness is arised from two distinct resources: the first part is by "switching" and the second part is by "diffusion". The former says that the switching itself follows a pattern of alternating periods as a Markov process. The latter says the there exist a certain amount of uncertainty for each dynamics by the mesocopic point of view. Based on those two parts, we can write a coupled Fokker-Plank equation to model the

Figure 7.4: Positions of the three overdamped particles. The red lines indicate systolic phases and the blue lines indicate diastolic phases. The dynamics of $\theta_1, \theta_2, \theta_3$ are monotonic in each phase $[t_i, t_{i+1})$ regardless of the pattern of alternating periods. Specifically, θ_3 is monotonically increasing for all t.

Figure 7.5: The red lines indicate systolic phases and the blue lines indicate diastolic phases. $\frac{d\theta_1}{dt}(t)$, $\frac{d\theta_2}{dt}(t)$ have discontinuity at each perturbation, particularly, the sign changes after each perturbation. On the other hand, $\frac{d\theta_3}{dt}(t)$ is a continuous function and it is positive for all t regardless of the pattern of alternating periods.

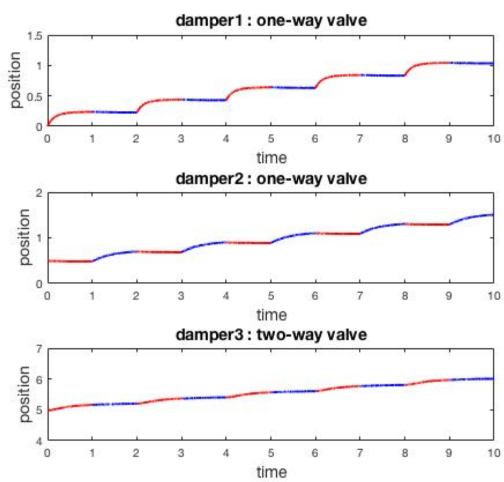

Figure 7.4: Positions of the three particles.

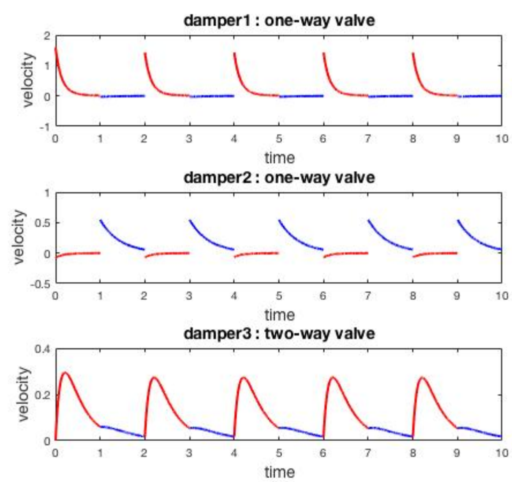

Figure 7.5: Velocities of the three particles.

cardiovascular system with noise. This model can be attacked by the stochastic analysis of coupled diffusion in a non-equilibrium system [6, 159]. It is also called "coupled diffusion process" [174]. The coupled diffusion process for our model Eqs. (7.4) and (7.5) are given by Eqs. (7.1) and (7.2) plus with noises, respectively (using shorthand notations here):

$$d\hat{\theta} = \mathbf{b}^{(d)}(\hat{\theta})dt + \mathbf{\Gamma}^{(d)}dW_t \tag{7.4}$$

$$d\hat{\theta} = \mathbf{b}^{(s)}(\hat{\theta})dt + \mathbf{\Gamma}^{(s)}dW_t \tag{7.5}$$

in which $\hat{\theta} = (\theta_1, \theta_2, \theta_3)$. The functions $\mathbf{b}_d, \mathbf{b}_s$ are for *drift* and $\mathbf{\Gamma}_d, \mathbf{\Gamma}_s$ are the coefficients of *diffusion* for the diastolic and systolic phase, respectively.

Step 4: Find probabilistic representations for dynamics (7.4) and (7.5). The corresponding Fokker-Plank equation for the diastolic phase / the systolic phase can be written as

$$\begin{aligned}
\frac{\partial P^{(d)}(\hat{\theta}, t)}{\partial t} &= -\sum_{i=1}^{3} \frac{\partial}{\partial \theta_i} J^{(d)}(\hat{\theta}, t) \\
&= \sum_{i,j=1}^{3} \frac{\partial^2}{\partial \theta_i \partial \theta_j} [D_{ij}^{(d)} P^{(d)}(\hat{\theta}, t)] - \sum_{i=1}^{3} \frac{\partial}{\partial \hat{\theta}} [b_i^{(d)}(\hat{\theta}) P^{(d)}(\hat{\theta}, t)], \\
\frac{\partial P^{(s)}(\hat{\theta}, t)}{\partial t} &= -\sum_{i=1}^{3} \frac{\partial}{\partial \theta_i} J^{(s)}(\hat{\theta}, t) \\
&= \sum_{i,j=1}^{3} \frac{\partial^2}{\partial \theta_i \partial \theta_j} [D_{ij}^{(s)} P^{(s)}(\hat{\theta}, t)] - \sum_{i=1}^{3} \frac{\partial}{\partial \hat{\theta}} [b_i^{(s)}(\hat{\theta}) P^{(s)}(\hat{\theta}, t)].
\end{aligned} \tag{7.6}$$

in which $\mathbf{D}^{(d)} = \mathbf{\Gamma}^{(d)} (\mathbf{\Gamma}^{(d)})^T/2$ and $\mathbf{D}^{(s)} = \mathbf{\Gamma}^{(s)} (\mathbf{\Gamma}^{(s)})^T/2$. On top of Eqs. (7.6), we need to couple them by the possibly $\hat{\theta}$-dependent rates when transitions occur between the states: $f^{(d)}(\hat{\theta})$ is the rate from diastolic state to the systolic state and $f^{(s)}(\hat{\theta})$ is the rate from systolic state to the diastolic state [155, 7]. Then the time-dependent probability densities follow a pair of partial differential equations

$$\begin{aligned}\frac{\partial P^{(d)}(\hat{\theta},t)}{\partial t} &= \sum_{i,j=1}^{3}\frac{\partial^2}{\partial\theta_i\partial\theta_j}[D_{ij}^{(d)}P^{(d)}(\hat{\theta},t)] - \sum_{i=1}^{3}\frac{\partial}{\partial\hat{\theta}}[b_i^{(d)}(\hat{\theta})P^{(d)}(\hat{\theta},t)] \\ &\quad - f^{(d)}(\hat{\theta})P^{(d)}(\hat{\theta},t) + f^{(s)}(\hat{\theta})P^{(d)}(\hat{\theta},t), \\ \frac{\partial P^{(s)}(\hat{\theta},t)}{\partial t} &= \sum_{i,j=1}^{3}\frac{\partial^2}{\partial\theta_i\partial\theta_j}[D_{ij}^{(s)}P^{(s)}(\hat{\theta},t)] - \sum_{i=1}^{3}\frac{\partial}{\partial\hat{\theta}}[b_i^{(s)}(\hat{\theta})P^{(s)}(\hat{\theta},t)] \\ &\quad - f^{(s)}(\hat{\theta})P^{(s)}(\hat{\theta},t) + f^{(d)}(\hat{\theta})P^{(s)}(\hat{\theta},t).\end{aligned} \tag{7.7}$$

This is known as the corresponding coupled Fokker-Planck equations to a coupled diffusion system in mathematics [174].

Step 5: To generalize the concept of entropy production rate to the coupled diffusion system, it has been shown with an additional term from the transitions among inner states (diastole/systole) such as

$$e_p = e_p^{(d)} + e_p^{(s)} + e_p^{(m)}, \tag{7.8}$$

where

$$e_p^{(d)} = \int_{\mathrm{R}^3}\mathbf{J}[P^{(d)}(\hat{\theta},t)](D)^{-1}\mathbf{J}[P^{(d)}(\hat{\theta},t)]P^{(d)}(\hat{\theta},t)^{-1}d\hat{\theta}, \tag{7.9}$$

$$e_p^{(s)} = \int_{\mathrm{R}^3}\mathbf{J}[P^{(s)}(\hat{\theta},t)](D)^{-1}\mathbf{J}[P^{(s)}(\hat{\theta},t)]P^{(s)}(\hat{\theta},t)^{-1}d\hat{\theta}, \tag{7.10}$$

$$e_p^{(m)} = \int_{\mathrm{R}^3}(f^{(d)}(\hat{\theta})P^{(d)}(\hat{\theta},t) - f^{(s)}(\hat{\theta})P^{(s)}(\hat{\theta},t))\log\left(\frac{f^{(d)}(\hat{\theta})P^{(d)}(\hat{\theta},t)}{f^{(s)}(\hat{\theta})P^{(s)}(\hat{\theta},t)}\right)d\hat{\theta}. \tag{7.11}$$

The $e_p^{(d)}$ and $e_p^{(s)}$ measures the non-equilibrium in the diastolic phase and in the systolic phase, separately; The additional $e_p^{(m)}$ measures the transitions among inner states. In physiology, the former represents the entropy produced by blood flows through resistance; the later represents the entropy produced by stochasticity as a consequence of the heart-rate fluctuations. Applied to pathophysiology, the former is abnormal when the resistance to blood flow is high (e.g. valvular stenosis); the later is abnormal when the heart rhythm is highly stochastic (e.g. cardiac arrhythmia).

7.2 Homeostasis in biology, relation to thermodynamics

Homeostasis is the tendency to resist variations in the external environment in order to maintain a stable internal environment [26]. Homeostasis has been widely modeled by control systems in literature and textbooks [105, 2]. Most of models can be simplified into a toy model with two compartments, *effective one* and *reserved one*, which are connected by two pathways: flow into and flow out of the effective compartment. Homeostasis typically involves *negative feedback loops*, in which signals from the effective compartment regulate the ratio of the inflow and the outflow in order to counteract the perturbation. However, if the ratio of two opposite flows is fixed, the total *cycle flux* composed by the two opposite flows is not obviously controlling homeostasis. Therefore, instead of using the ratio of the two flows, we find another way to quantify the strength of cycle flux and relate it to the regulation of homeostasis by the following three steps: First, we were inspired by the *linear, unimolecular single-molecule enzyme kinetics, phosphorylation-dephosphorylation cycle* (PdPc) model [167] with negative feedback controls. Second, since this enzyme model (PdPc) and the homeostatic models share the same mathematical framework, we can apply "free energy" which is a well-defined quantity of cycle flux in the PdPc to the homeostatic systems. In the homeostatic systems, this energy is corresponding to the driving force for the corresponding flow. Third, we show that the negative feedback loops and the free energy of cycle flux are inseparably interconnected.

From the above, we mathematically prove that free energy released by ATP hydrolysis in the cycle amplifies the sensitivities of negative feedback controls, so the homeostatic systems are more robust in a nonequilibrium steady state (NESS). Our results provide an explanation for physiology existing several "non-effective cycles" which run opposite pathways simultaneously and cost huge amount of energy: for instances, 1. *Futile cycles* in glycosis and gluconeogenesis [104] [166] . 2. *Fluid and electrolytes cycles* in renal filtration and absorption [53]. The former is important for glucose homeostasis and the later is for body fluid homeostasis. Both are regulated by hormone negative feedback loops. Here is a new perspective of homeostasis in human physiology: sufficient available energy input in the cycles are necessary to maintain the sensitivities of hormone

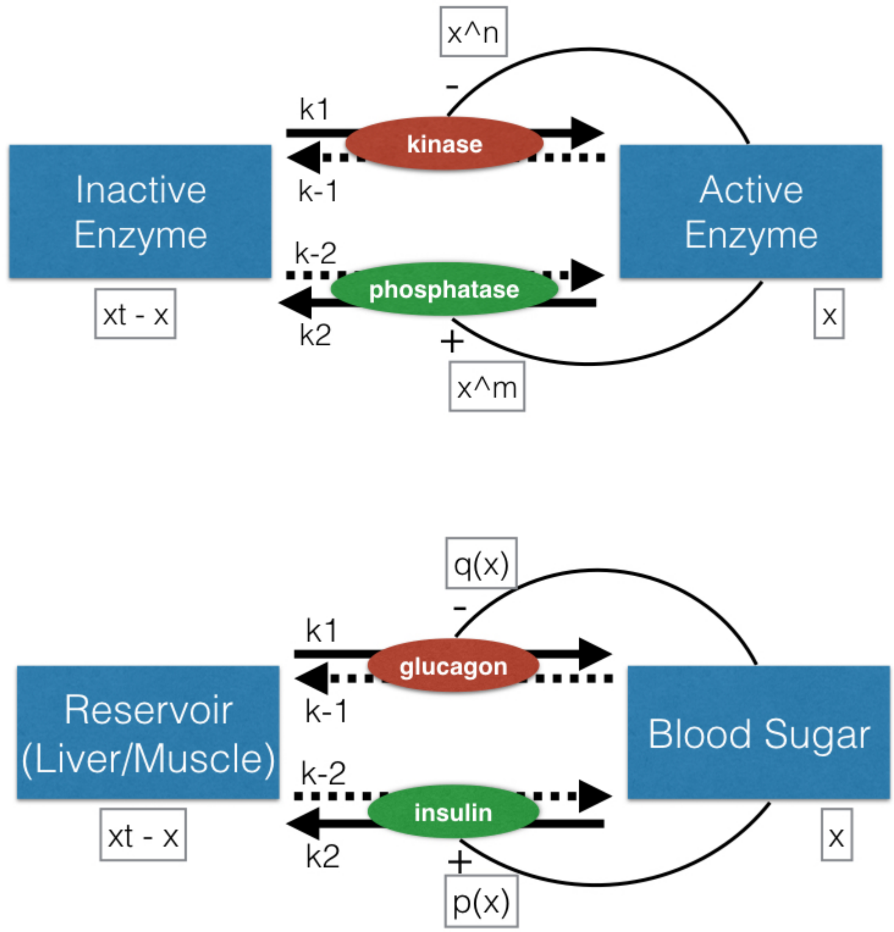

Figure 7.6: Negative feedback loops in enzyme reactoins and blood glucose control.

regulations. Furthermore, based on our findings, we are able to relate diseases of homeostasis to energy insufficiency and therefore provide different strategies in clinical treatments. For example, mitochondrial dysfunction has been widely discussed in Type-2 diabetes [127, 139].

Nonequilibrium steady state of a biochemical system: the enzyme reactions in open systems — For biochemical open systems, here is an example of homeostasis in a living cell with the phosphorylation-dephosphorylation cycle (PdPc) of an enzyme. In this model, a substrate enzyme E, a kinase K, and a phosphatase P, are the three key elements. Based on those three elements, PdPc consists of two chemical reactions:

$$\mathrm{ATP} + \mathrm{E} \xrightarrow{k_1} \mathrm{ADP} + \mathrm{E}^*, \tag{7.12}$$

$$\mathrm{E}^* \xrightarrow{k_2} \mathrm{E} + \mathrm{P}_i, \tag{7.13}$$

in which the reaction (7.12) represents that the phosphorylation of the substrate protein E is catalyzed by kinase K, and the reaction (7.13) represents that the dephosphorylation is catalyzed by phosphatase P. In addition, kinase K and phosphatase P can transform between an inactive state and an active state. In the case that the conversion of kinase from the inactive form to the active form involves the binding of E*, this is called *autocatalysis* and has been carefully studied by Bishop and Qian [167, 18]. Inspired by Bishop and Qian, we here focus on the conversion of phosphatase from the inactive to active form by the binding of E* with different order m

$$\mathrm{P} + \mathrm{mE}^* \overset{\mathrm{K}_a}{\rightleftharpoons} \mathrm{P}^*. \tag{7.14}$$

And we assume the reaction (7.14) is rapid, so the active phosphatase concentration reaches quasi-steady state $[\mathrm{P}^*] = \mathrm{K}_a [\mathrm{P}][\mathrm{E}^*]^m$. Applying the quasi-steady state Eq. (7.14) to the PdPC given in the pair of reactions (7.12) and (7.13), we then obtain *homeostasis* in the enzyme reaction: increasing P* is a consequence of more E*; however, P* further converts E* back to E by the reaction (7.13) hence less E*. This entire story of homeostasis can be achieve by a parallel mechanism of the conversion of kinase from the active form to then inactive form. The regulation of enzyme homeostasis has to be controlled by a negative feedback loop either activating the backward reaction ($\mathrm{E}^* \longrightarrow \mathrm{E}$) or deactivating the forward reaction ($\mathrm{E} \longrightarrow \mathrm{E}^*$). The above model is illustrated in Figure 7.6.

To transform the above chemical model to a kinetic scheme, we then have a differential equation:

$$\frac{dx}{dt} = \alpha(x_t - x) - \beta x - \epsilon x^{m+1} + \delta x^m (x_t - x), \tag{7.15}$$

in which variable x represents E*, x_t is a parameter for the total [E] + [E*], and

$$\alpha = k_1[\mathrm{ATP}][\mathrm{K}], \quad \beta = k_{-1}[\mathrm{ADP}][\mathrm{K}], \quad \epsilon = k_2 K_a[\mathrm{P}], \quad \delta = k_{-2}K_a[\mathrm{P}][\mathrm{P}_i] \tag{7.16}$$

are the rates for the reversible phosphorylation (α and β) and the reversible dephosphorylation (ϵ and δ) and the order m represents the strength of the negative feedback loop by (7.14). Then we introduce $\theta = \frac{[K]}{[P]}$ as a controlling parameter for the homeostasis. Let $m = 2$, as the system reaching steady states for Eq. (7.15),

$$\theta = \frac{[K]}{[P]} = \frac{\gamma\delta x^3 - \mu\delta x^2(x_t - x)}{\mu^2(x_t - x) - \mu x}, \tag{7.17}$$

in which

$$\gamma = \frac{k_1 k_2[\text{ATP}]}{k_{-1}k_{-2}[\text{ADP}][\text{P}_i]}, \quad \mu = \frac{k_1}{k_{-1}}, \quad \delta = \frac{k_{-2}}{k_{-1}}.$$

In particular, if the ATP hydrolysis is at the equilibrium, i.e., $\gamma = 1$, then

$$x = \left(\frac{\mu}{\gamma + \mu}\right) x_t,$$

which gives rise to a very special result that the steady state of active enzyme [E*] is independent of [K] and [P], which means the system is perfectly robust with respect to any perturbation from the environment (We here assume only [K] and [P] can be regulated by the environment).

Furthermore, we are also interested in the *sensitivity of homeostasis* with respect to θ and the *error of homeostasis*. First, we define sensitivity and error of homeostasis as

$$\text{Sensitivity} = \frac{dx/x}{d\theta/\theta} \tag{7.18}$$

$$\text{Error} = \frac{x - x_0}{x_0} \approx \left.\frac{dx/x}{d\theta}\right|_{x_0} (\Delta\theta) \tag{7.19}$$

By our simulations (See Figures 7.7 & 7.8), we found that the sensitivity of homeostasis is increasing but the error of homeostasis is decreasing when we increase the value of γ. Using the sensitivity of homeostasis (7.18) to define the robustness of a system can be found Jia model [100] and applying error of homeostasis (7.19) to define the robustness of a system was introduced in Tu model [118]. In our results, adopting those two different definitions of homeostasis will give rise to opposite trends with respect to ATP hydrolysis (the ratio of ATP to ADP). This paradox requires further investigations.

Nonequilibrium steady state of physiological systems: blood glucose control and blood volume control —

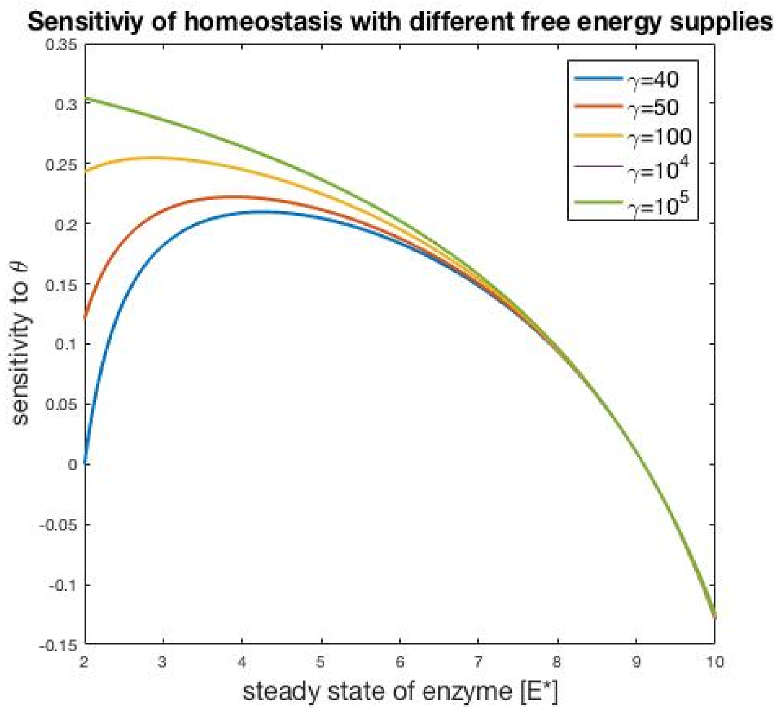

Figure 7.7: Sensitivity of homeostasis.

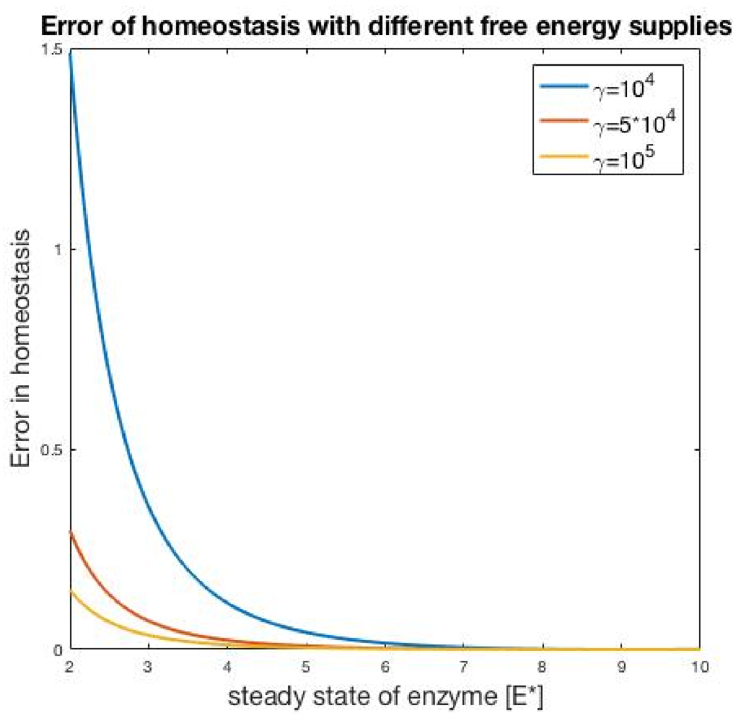

Figure 7.8: Error in homeostasis.

Let us illustrate a model of homeostasis in human physiology by analogy with the enzyme kinetic model, see Figure 7.6. We divide a system into two parts: the central part and the reservoir. Consider an example of human body sugar model, the central part is corresponding to the blood sugar and the reservoir is corresponding to the sugar storage in the liver and tissues; Consider another example of blood volume modeling, the central part is for central blood volume (brain, heart, and lung) and the reservoir is for peripheral blood volume (skin and muscle). Apply this idea back to our enzyme models, the former is exactly active enzyme $[\mathrm{E}^*]$ and the later should be inactive enzyme $[\mathrm{E}]$.

Therefore, we can model negative feedback loops as this way: increasing blood sugar level activates insulin or inhibits glucagon in order to achieve homeostasis. Same as the blood volume modeling: increasing central blood volume actives the parasympathetic system or inhibits the sympathetic system. Activation or inhibition is indicated by the function $p(x, \theta)$ or $q(x, \theta)$, respectively. The choice of specific function forms p, q depends on physiological experiments or insights. In particular, applying back to our enzyme models, a specific activation function is given by $p(x, \theta) = [\mathrm{P}(\theta)]\mathrm{x}^n$, where $[\mathrm{P}(\theta)]$ is the amount of preactived phosphatase (depending on an environmental factor θ), and the preactived phosphatase will be activated through binding the active enzyme with order n.

Here, we introduce a rather general model for negative feedback controls in physiology

$$\frac{dx}{dt} = a\frac{(x_t - x)}{q(\theta, x)} - b\frac{x}{q(\theta, x)} - cp(\theta, x)x + dp(\theta, x)(x_t - x), \tag{7.20}$$

in whcih

$$a = k_1 K_a, \quad b = k_{-1} K_a, \quad c = k_2 K_i, \quad d = k_{-2} K_i. \tag{7.21}$$

For steady states analysis, we solve Eq. (7.20) with grouping parameters (assume functions are seperable $p(\theta, x) = p_1(x)p_2(\theta)$ and $q(\theta, x) = q_1(x)q_2(\theta)$), then we obtain

$$\Theta(\theta) = \frac{\frac{a}{d}(-(1 + \frac{b}{a})x + x_t)}{(p_1(x)q_1(x))((1 + \frac{b}{a}\gamma)x - x_t)}, \tag{7.22}$$

where function $\Theta(\theta) = \frac{p_2(\theta)}{q_2(\theta)}$ is dependent on the environmental factor θ and $\gamma = \frac{ac}{bd}$ represents the total energy flux in the system.

Based on the above state-state analysis, we are further interested in the robustness of system with respect to different "strength" of activation and inhibition in the negative feedback loops which are characterized by the function $p_1(x)$ or $p_2(x)$. Additionally, we are also interested in the robustness of system with respect to different "energy flux" γ. Therefore, we derive the sensitivity of x to θ, which has been defined above,

$$\left|\frac{d\ln x}{d\ln\Theta}\right| = \frac{\frac{1}{x}}{\left(\frac{\ln(p_1(x)q_1(x))}{dx} + \frac{\frac{b}{a}(\gamma-1)x_t}{((\frac{b}{a}\gamma+1)x-x_t)(-(1+\frac{b}{a})x+x_t)}\right)}. \tag{7.23}$$

Particularly, in the enzyme models, the first term in denominator is $\frac{d\ln(p_1(x)q_1(x))}{dx} = \frac{d\ln(x^{n+m})}{dx} = \frac{n+m}{x}$. Therefore, if we fix $x, \Theta, \frac{b}{a}$, and γ, then the sensitivity will decrease with higher $n+m$ (Figure 7.9). The second term in denominator is non-positive and monotonic increasing with higher γ. Therefore, if we fix $x, \Theta, \frac{b}{a}$, and $p_1(x)q_1(x)$, then the sensitivity will increase with higher γ (Figure 7.10).

So far, we assume only activation and inhibition functions depend on environments. In physiology or cell biology, x_t (the total amount of sugar level, blood volume and enzyme) can be influenced by fluctuating environments as well. So we are also interested in how robust of x is with respect to x_t. So we define another sensitivity to x_t here:

$$\left|\frac{d\ln x}{d\ln x_t}\right| = \frac{\frac{x_t}{x}}{\frac{x_t}{x} + \left(\frac{\ln(p_1(x)q_1(x))}{dx}\right)\left(\frac{((\frac{b}{a}\gamma+1)x-x_t)(-(1+\frac{b}{a})x+x_t)}{\frac{b}{a}(\gamma-1)x}\right)}. \tag{7.24}$$

For the enzyme models, $\frac{d\ln(p_1(x)q_1(x))}{dx} = \frac{d\ln(x^{n+m})}{dx} = \frac{n+m}{x}$. Therefore, if we fix $x, \Theta, \frac{b}{a}$, and γ, then the sensitivity will decrease with higher $n+m$ (Figure 7.11). And the term depending on γ is nonnegative and monotonic increasing with higher γ. Therefore, if we fix $x, \Theta, \frac{b}{a}$, and $p_1(x)q_1(x)$, then the sensitivity will decrease with higher γ (Figure 7.12). To further analyze and compare the results in Figures 7.9, 7.10, 7.11, and 7.12 with physical interpretations will be at the core of the next step in this project in the future.

Three strategies for the survival in fluctuating environments— In fluctuating environments, organisms exhibit different patterns of their behavior in order to survive and maintain population growth rate. These patterns are often called "evolutionary strategies" following the game-theoretical representations of Darwinian evolution theory. *Homeostasis* and *acclimation* are two

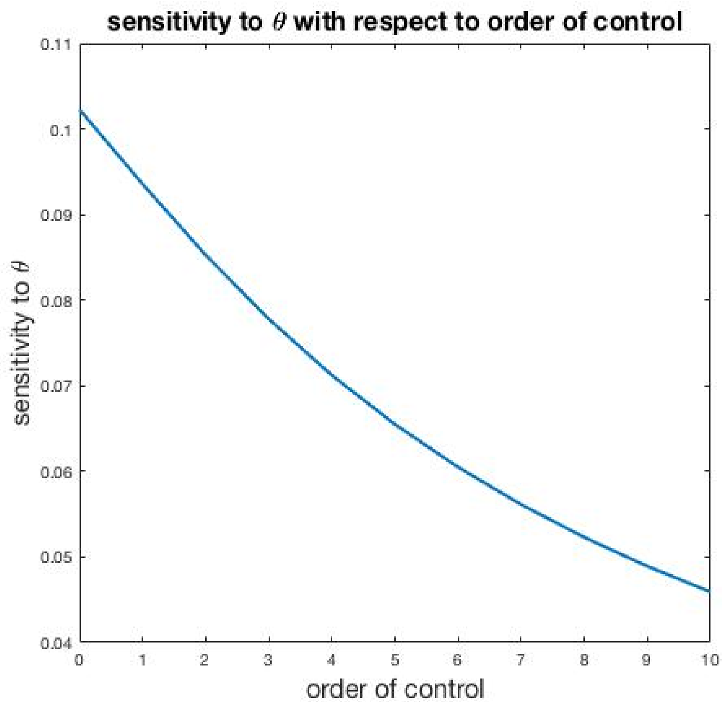

Figure 7.9: Sensitivity to θ decreases with higher order controls $n + m$.

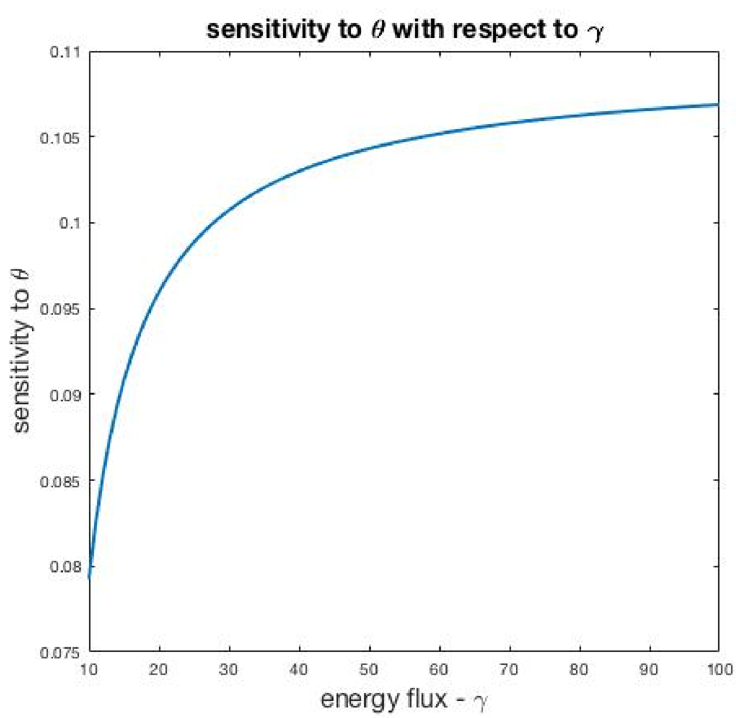

Figure 7.10: Sensitivity to θ increases with higher amount of free energy γ.

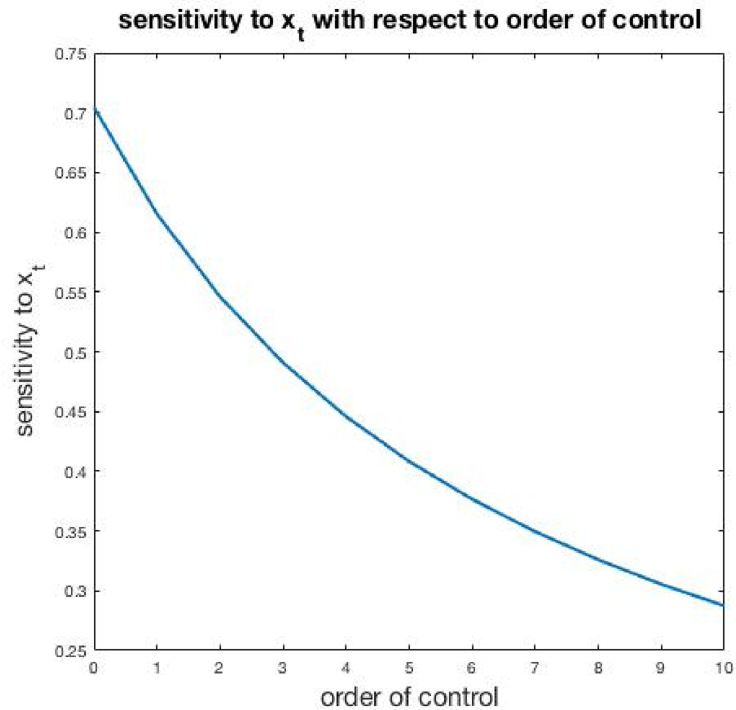

Figure 7.11: Sensitivity to x_t decreases with higher order controls $n + m$.

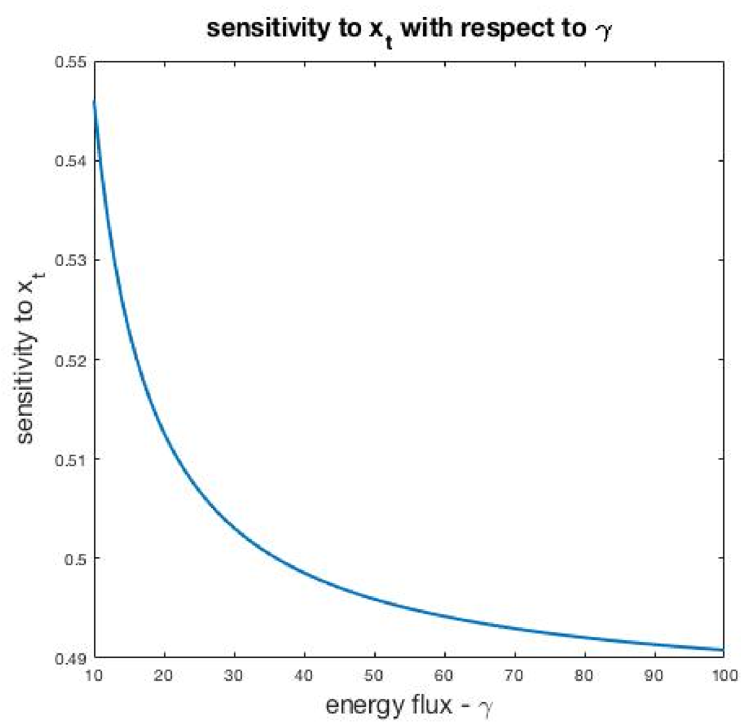

Figure 7.12: Sensitivity to x_t decreases with higher amount of free energy γ.

common “strategies”, e.g., mechanisms, by which an individual organism to live robustly in fluctuating environments; note the emphasize here is *individual*. Biologically, homeostasis and acclimation are two seemingly opposite ways to overcome environmental changes. The former is about stabilizing but the later is about changing the internal state of an organism when facing varying external environments. We will show that, however, mathematically these two concepts actually arise from a single dynamical system with two types of state variables. With respect to varying environments, the first type is called *homeostatic* while the second type *acclimative*.

In terms of this mathematical representation, we discover that homeostasis and acclimation are actually two sides of the same coin. We extensively explore an canonical form of this model by introducing (i) a measure for the “robustness” of the homeostatic variable and (ii) a notion of “nonequilibirum energy cost”. According to the model, an individual is actually a *quasi-species* in an epigenetic landscape according Eigen-Schuster [54], and its apparent growth rate provides the essential connection between the individual’s behaviors (homeostasis and acclimation) and the behavior of a population of i.i.d. individuals, such as *adaptation*. Here we use the term adaptation in its original Darwinian sense, e.g., the tendency of a population of organisms becoming “fitter” to its environment, enhancing their evolutionary fitness [186]. From the perspective of the evolution theory, both of homeostasis and acclimation can be developed through adaptation. One of the future goals in this project is to show that those three biologically different concepts - homeostasis, acclimation, and adaptation - eventually can be integrated into a unified mathematical model.

BIBLIOGRAPHY

[1] D.H. Ackley, G.E. Hinton, and R.J. Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive science*, 9(1):147–169, 1985.

[2] U. Alon. *An introduction to systems biology: design principles of biological circuits*. CRC press, 2019.

[3] P.W. Anderson. More is different. *Science*, 177(4047):393–396, 1972.

[4] P. Ao. Potential in stochastic differential equations: novel construction. *Journal of physics A: mathematical and general*, 37(3):L25, 2004.

[5] P. Ao. Emerging of stochastic dynamical equalities and steady state thermodynamics from darwinian dynamics. *Communications in theoretical physics*, 49(5):1073, 2008.

[6] R.D. Astumian. Thermodynamics and kinetics of a brownian motor. *science*, 276(5314):917–922, 1997.

[7] R.D. Astumian and M. Bier. Mechanochemical coupling of the motion of molecular motors to atp hydrolysis. *Biophysical journal*, 70(2):637–653, 1996.

[8] R. Balian. *From Microphysics to Macrophysics: Methods and Applications of Statistical Physics. Volume II*. Springer Science & Business Media, 2007.

[9] J.R. Banavar, A. Maritan, and I. Volkov. Applications of the principle of maximum entropy: from physics to ecology. *Journal of Physics: Condensed Matter*, 22(6):063101, 2010.

[10] M. Barlow, K. Burdzy, and Á. Timár. Comparison of quenched and annealed invariance principles for random conductance model. *Probability Theory and Related Fields*, 164(3-4):741–770, 2016.

[11] A. Ben-Naim. Mixing and assimilation in systems of interaction particles. *American Journal of Physics*, 55(12):1105–1109, 1987.

[12] V. Benci and T. D'Aprile. The semiclassical limit of the nonlinear schrödinger equation in a radial potential. *Journal of Differential Equations*, 184(1):109–138, 2002.

[13] K.M. Bender and S.A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic methods and perturbation theory*. Springer Science & Business Media, 2013.

[14] H.C. Berg. *Random Walks in Biology*. Princeton University Press, 1993.

[15] P.G. Bergmann and J.L. Lebowitz. New approach to nonequilibrium processes. *Physical Review*, 99(2):578, 1955.

[16] W. Bialek. *Biophysics: Searching for Principles*. Princeton University Press, 2012.

[17] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

[18] L.M. Bishop and H. Qian. Stochastic bistability and bifurcation in a mesoscopic signaling system with autocatalytic kinase. *Biophysical journal*, 98(1):1–11, 2010.

[19] S. Bittanti, P. Bolzern, and P. Colaneri. Stability analysis of linear periodic systems via the lyapunov equation. *IFAC Proceedings Volumes*, 17(2):213–216, 1984.

[20] N. Bleistein and R.A. Handelsman. *Asymptotic Expansions of Integrals*. Courier Corporation, 1986.

[21] P.C. Bressloff and J.N. MacLaurin. A variational method for analyzing stochastic limit cycle oscillators. *SIAM Journal on Applied Dynamical Systems*, 17(3):2205–2233, 2018.

[22] W. Bryc. A remark on the connection between the large deviation principle and the central limit theorem. *Statistics & probability letters*, 18(4):253–256, 1993.

[23] W. Bryc and A. Dembo. Large deviations for quadratic functionals of gaussian processes. *Journal of Theoretical Probability*, 10(2):307–332, 1997.

[24] G.L. Butterfoss and J. Hermans. Boltzmann-type distribution of side-chain conformation in proteins. *Protein Science*, 12(12):2719–2731, 2003.

[25] M. Campisi. On the mechanical foundations of thermodynamics: The generalized helmholtz theorem. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 36(2):275–290, 2005.

[26] W.B. Cannon. Organization for physiological homeostasis. *Physiological reviews*, 9(3):399–431, 1929.

[27] M.E. Cates and V.N. Manoharan. Celebrating soft matter's 10th anniversary: Testing the foundations of classical entropy: colloid experiments. *Soft Matter*, 11(33):6538–6546, 2015.

[28] R. Chang and K.A. Goldsby. *Chemistry, 11th ed.* McGraw-Hill, New York., 2012.

[29] Y. Chen and S. Chen. Existence of positive definite solution to periodic riccati differential equation. *Proceedings of the 3rd World Congress on Intelligent Control and Automation (Cat. No. 00EX393)*, 4:2820–2823, 2000.

[30] Y.-C. Cheng and H. Qian. Stochastic limit-cycle oscillations of a nonlinear system under random perturbations. *Journal of Statistical Physics*, 182(3):1–33, 2021.

[31] Y.-C. Cheng, H. Qian, and Y. Zhu. Asymptotic behavior of a sequence of conditional probability distributions and the canonical ensemble. *Annales Henri Poincaré*, 22:1561–1627, 2021.

[32] Y.-C. Cheng, W. Wang, and H. Qian. Generalizing gibbsian statistical ensemble theory for strongly coupled heterogeneous systems. *arXiv preprint arXiv:1811.11321*, 2021.

[33] S. Chibbaro, L. Rondoni, and A. Vulpiani. *Reductionism, emergence and levels of reality*. Springer, New York, 2014.

[34] T.M. Cover and J.A. Thomas. Information theory and statistics. *Elements of Information Theory*, 1(1):279–335, 1991.

[35] H. Cramér. A century with probability theory: Some personal recollections. *The annals of probability*, 4(4):509–546, 1976.

[36] L. Cronin and S.I. Walker. Beyond prebiotic chemistry. *Science*, 352(6290):1174–1175, 2016.

[37] A. Dembo and O. Zeitouni. Refinements of the Gibbs conditioning principle. *Probability theory and related fields*, 104(1):1–14, 1996.

[38] A Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer-Verlag Berlin Heidelberg, 2010.

[39] J.D. Deuschel, D.W. Stroock, and H. Zessin. Microcanonical distributions for lattice gases. *Communications in mathematical physics*, 139(1):83–101, 1991.

[40] K.A. Dill and S. Bromberg. *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience*. Garland Science, 2012.

[41] R. Dobbertin. On functional relations between reduced distribution functions and entropy production by non-hamiltonian perturbations. *Physica Scripta*, 14(3):85, 1976.

[42] R.L. Dobrushin and B. Tirozzi. The central limit theorem and the problem of equivalence of ensembles. *Communications in Mathematical Physics*, 54(2):173–192, 1977.

[43] M.D Donsker. *An Invariance Principle for Certain Probability Limit Theorems*. AMS, 1951.

[44] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain markov process expectations for large time, i. *Communications on Pure and Applied Mathematics*, 28(1):1–47, 1975.

[45] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain markov process expectations for large time, ii. *Communications on Pure and Applied Mathematics*, 28(2):279–301, 1975.

[46] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain markov process expectations for large time—iii. *Communications on pure and applied Mathematics*, 29(4):389–461, 1976.

[47] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. IV. *Communications on Pure and Applied Mathematics*, 36(2):183–212, 1983.

[48] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain markov process expectations for large time. iv. *Communications on Pure and Applied Mathematics*, 36(2):183–212, 1983.

[49] M.D. Donsker and S.R.S. Varadhan. Large deviations for stationary gaussian processes. *Communications in Mathematical Physics*, 97(1-2):187–210, 1985.

[50] R. Durrett. *Probability: Theory and Examples*, volume 49. Cambridge university press, 2019.

[51] M. Dykman, X. Chu, and J. Ross. Stationary probability distribution near stable limit cycles far from hopf bifurcation points. *Physical Review E*, 48(3):1646, 1993.

[52] M.I. Dykman and M.A. Krivoglaz. Theory of fluctuational transitions between stable states of a nonlinear oscillator. *Sov. Phys. JETP*, 50(1):30–37, 1979.

[53] D.C. Eaton and J.P. Pooler. *Vander's renal physiology*. Mc Graw Hil Medicall, 2009.

[54] M. Eigen, J. McCaskill, and P. Schuster. Molecular quasi-species. *The Journal of Physical Chemistry*, 92(24):6881–6891, 1988.

[55] M. Esposito and C. Van den Broeck. Three detailed fluctuation theorems. *Physical review letters*, 104(9):090601, 2010.

[56] L.C. Evans. Partial differential equations. graduate studies in mathematics. *American mathematical society*, 2:1998, 1998.

[57] L.C. Evans and H. Ishii. A PDE approach to some asymptotic problems concerning random differential equations with small noise intensities. *Annales de l'Institut Henri Poincare (C) Non Linear Analysis*, 2(1):1–20, 1985.

[58] W. Feller. The general diffusion operator and positivity preserving semi-groups in one dimension. *Annals of Mathematics*, pages 417–436, 1954.

[59] J.-M. Fellous, P.H.E. Tiesinga, P.J. Thomas, and T.J. Sejnowski. Discovering spike patterns in neuronal responses. *Journal of Neuroscience*, 24(12):2989–3001, 2004.

[60] H. Feng, B. Han, and J. Wang. Landscape and global stability of nonadiabatic and adiabatic oscillations in a gene network. *Biophysical journal*, 102(5):1001–1010, 2012.

[61] H. Feng, K. Zhang, and J. Wang. Non-equilibrium transition state rate theory. *Chemical Science*, 5(10):3761–3769, 2014.

[62] J. Feng and T.G. Kurtz. *Large Deviations for Stochastic Processes*, volume 131. American Mathematical Soc., 2006.

[63] W.H. Fleming. Exit probabilities and optimal stochastic control. *Applied Mathematics and Optimization*, 4(1):329–346, 1977.

[64] H. Frauenfelder and P.G. Wolynes. Biomolecules: where the physics of complexity and simplicity meet. *Physics Today;(United States)*, 47(2), 1994.

[65] M.I. Freidlin and A.D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer, 1998.

[66] A. Friedli and Y. Velenik. *Statistical mechanics of lattice systems: a concrete mathematical introduction*. Cambridge University Press, 2017.

[67] G. Gallavotti. *Statistical mechanics: A short treatise*. Springer, Berlin, 1999.

[68] H. Gang. Lyapounov function and stationary probability distributions. *Zeitschrift für Physik B Condensed Matter*, 65(1):103–106, 1986.

[69] C. Gardiner. *Stochastic Methods*. Springer Berlin, 2009.

[70] H. Ge and H. Qian. Physical origins of entropy production, free energy dissipation, and their mathematical representations. *Physical Review E*, 81(5):051133, 2010.

[71] H. Ge and H. Qian. Maximum entropy principle, equal probability a priori and gibbs paradox. *arXiv preprint arXiv:1105.4118*, 2011.

[72] H. Ge and H. Qian. Analytical mechanics in stochastic dynamics: most probable path, large-deviation rate function and hamilton–jacobi equation. *International Journal of Modern Physics B*, 26(24):1230012, 2012.

[73] H. Ge and H. Qian. Landscapes of non-gradient dynamics without detailed balance: Stable limit cycles and multiple attractors. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(2):023140, 2012.

[74] H. Ge and H. Qian. Mesoscopic kinetic basis of macroscopic chemical thermodynamics: A mathematical theory. *Physical Review E*, 94(5):052150, 2016.

[75] H. Ge and H. Qian. Mathematical formalism of nonequilibrium thermodynamics for nonlinear chemical reaction systems with general rate law. *Journal of Statistical Physics*, 166(1):190–209, 2017.

[76] H. Ge, M. Qian, and H. Qian. Stochastic theory of nonequilibrium steady states. part ii: Applications in chemical biophysics. *Physics Reports*, 510(3):87–118, 2012.

[77] H.O. Georgii. The equivalence of ensembles for classical systems of particles. *Journal of statistical physics*, 80(5-6):1341–1378, 1995.

[78] H.O. Georgii. *Gibbs Measures and Phase Transitions*. De Gruyter studies in mathematics. De Gruyter, 2011.

[79] Z. Ghahramani. An introduction to hidden markov models and bayesian networks. *Int. J. Pattern Recogn.*, 15:9–42, 2001.

[80] J.W. Gibbs. *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to The Rational Foundation of Thermodynamics*. C. Scribner's sons, 1902.

[81] I.I. Gikhman and A.V. Skorokhod. *Stochastic Differential Equations [by] I.I. Gihman [and] A.V. Skorohod*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer-Verlag, 1972.

[82] R. Graham and H. Haken. Generalized thermodynamic potential for markoff systems in detailed balance and far from thermal equilibrium. *Zeitschrift für Physik A Hadrons and nuclei*, 243(3):289–302, 1971.

[83] R. Graham and T. Tél. On the weak-noise limit of fokker-planck models. *Journal of statistical physics*, 35(5):729–748, 1984.

[84] R. Graham and T. Tél. Weak-noise limit of fokker-planck models and nondifferentiable potentials for dissipative dynamical systems. *Physical Review A*, 31(2):1109, 1985.

[85] E. Grenier. Semiclassical limit of the nonlinear schrödinger equation in small time. *Proceedings of the American Mathematical Society*, 126(2):523–530, 1998.

[86] M. Haw. *Middle World: The Restless Heart of Matter and Life*. Springer, 2016.

[87] L.T. Hill. *Statistical Mechanics: Principles and Selected Applications*. Dover Publications, 1987.

[88] T.L. Hill. *An introduction to statistical thermodynamics*. Dover, New York, 1986.

[89] C.J. Holland. Stochastically perturbed limit cycles. *Journal of Applied Probability*, 15(2):311–320, 1978.

[90] J.J. Hopfield. Physics, computation, and why biology looks so different. *Journal of Theoretical Biology*, 171(1):53–60, 1994.

[91] G. Hu. Stationary solution of master equations in the large-system-size limit. *Physical Review A*, 36(12):5782, 1987.

[92] K. Huang. *Statistical Mechanics*. Wiley Eastern, 1975.

[93] S. Huang, I. Ernberg, and S. Kauffman. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin. Cell Dev. Biol.*, 20(7):869–876, 2009.

[94] Y. Imry. *Introduction to Mesoscopic Physics, Second Edition*. Oxford University Press, 2002.

[95] J. Jacod and A.N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer, Berlin, Heidelberg, 1987.

[96] C. Jarzynski. Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale. *Annu. Rev. Condens. Matter Phys.*, 2(1):329–351, 2011.

[97] C. Jarzynski. Stochastic and macroscopic thermodynamics of strongly coupled systems. *Physical Review X*, 7(1):011008, 2017.

[98] E.T. Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[99] E.T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge university press, 2003.

[100] C. Jia, M. Qian, Y. Kang, and D. Jiang. Modeling stochastic phenotype switching and bet-hedging in bacteria: stochastic nonlinear dynamics and critical state identification. *Quantitative Biology*, 2(3):110–125, 2014.

[101] D.-Q. Jiang, M. Qian, and M.-P. Qian. *Mathematical theory of nonequilibrium steady states: on the frontier of probability and dynamical systems*. Springer, Berlin, Heidelberg, 2004.

[102] M. Kac. *Probability and related topics in physical sciences*, volume 1. American Mathematical Soc., 1959.

[103] M. Kardar. *Statistical Physics of Fields*. Cambridge University Press, 2007.

[104] J. Katz and R. Rognstad. Futile cycling in glucose metabolism. *Trends in Biochemical Sciences*, 3(3):171–174, 1978.

[105] J.P. Keener and J. Sneyd. *Mathematical physiology*, volume 1. Springer, 1998.

[106] J. Keizer. Nonequilibrium thermodynamics and the stability of states far from equilibrium. *Accounts of Chemical Research*, 12(7):243–249, 1979.

[107] J. Keizer. *Statistical Thermodynamics of Nonequilibrium Processes*. Springer-Verlag, 1987.

[108] A.I. Khinchin. *Mathematical Foundations of Statistical Mechanics*. Courier Corporation, 1949.

[109] A.I. Khinchin. *Mathematical foundations of information theory*. Dover, New York, 1957.

[110] J.G. Kirkwood. Statistical mechanics of fluid mixtures. *The Journal of chemical physics*, 3(5):300–313, 1935.

[111] H.A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.

[112] R. Kubo, K. Matsuo, and K. Kitahara. Fluctuation and relaxation of macrovariables. *Journal of Statistical Physics*, 9(1):51–96, 1973.

[113] S. Kullback. *Information Theory and Statistics*. Courier Corporation, 1997.

[114] C. Kurrer and K. Schulten. Effect of noise and perturbations on limit cycle systems. *Physica D: Nonlinear Phenomena*, 50(3):311–320, 1991.

[115] G.T. Kurtz. The central limit theorem for markov chains. *The Annals of Probability*, pages 557–560, 1981.

[116] T.G. Kurtz. The relationship between stochastic and deterministic models for chemical reactions. *The Journal of Chemical Physics*, 57(7):2976–2978, 1972.

[117] C. Kwon, P. Ao, and D.J. Thouless. Structure of stochastic dynamics near fixed points. *Proceedings of the National Academy of Sciences*, 102(37):13029–13033, 2005.

[118] G. Lan, P. Sartori, S. Neumann, V. Sourjik, and Y. Tu. The energy–speed–accuracy trade-off in sensory adaptation. *Nature physics*, 8(5):422–428, 2012.

[119] L.D. Landau and E.M. Lifshitz. *Statistical Physics*. Pergamon Press, Oxford, 1958.

[120] L.D. Landau and E.M. Lifshitz. *Statistical Physics (Course of Theoretical Physics vol 5)*. Pergamon Oxford, 1958.

[121] O.E. Lanford. Entropy and equilibrium states in classical statistical mechanics. In *Statistical Mechanics and Mathematical Problems*, pages 1–113. Springer, 1973.

[122] M. Levitt. The birth of computational structural biology. *Nature structural biology*, 8(5):392–393, 2001.

[123] J.T. Lewis, C.E. Pfister, and W.G. Sullivan. Entropy, concentration of probability and conditional limit theorems. *Markov Process. Related Fields*, 1(3):319–386, 1995.

[124] C. Li and J. Wang. Landscape and flux reveal a new global view and physical quantification of mammalian cell cycle. *Proceedings of the National Academy of Sciences*, 111(39):14130–14135, 2014.

[125] T.M. Liggett. *Interacting particle systems*. Springer-Verlag, New York, 1985.

[126] L. Lin, H. Yu, and X. Zhou. Quasi-potential calculation and minimum action method for limit cycle. *Journal of Nonlinear Science*, 29(3):961–991, 2019.

[127] B.B. Lowell and G.I. Shulman. Mitochondrial dysfunction and type 2 diabetes. *Science*, 307(5708):384–387, 2005.

[128] Z. Lu and H. Qian. Emergence and breaking of duality symmetry in thermodynamic behavior: Repeated measurements and macroscopic limit. *arXiv preprint arXiv:2009.12644*, 2020.

[129] Y.-A. Ma and H. Qian. A thermodynamic theory of ecology: Helmholtz theorem for lotka–volterra equation, extended conservation law, and stochastic predator–prey dynamics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2183):20150456, 2015.

[130] Y.-A. Ma, Q. Tan, R. Yuan, B. Yuan, and P. Ao. Potential function in a continuous dissipative chaotic system: Decomposition scheme and role of strange attractor. *International Journal of Bifurcation and Chaos*, 24(02):1450015, 2014.

[131] M.C. Mackey. The dynamic origin of increasing entropy. *Reviews of Modern Physics*, 61(4):981, 1989.

[132] M.L. Marcos and J. Echave. Too packed to change: side-chain packing and site-specific substitution rates in protein evolution. *PeerJ*, 3:e911, 2015.

[133] A. Martin-Löf. The equivalence of ensembles and the gibbs phase rule for classical lattice systems. *Journal of Statistical Physics*, 20(5):557–569, 1979.

[134] A. Martin-Löf. *Statistical mechanics and the foundations of thermodynamics*. Springer-Verlag Berlin Heidelberg, 1979.

[135] M. Mendler, J. Falk, and B. Drossel. Analysis of stochastic bifurcations with phase portraits. *PloS one*, 13(4), 2018.

[136] Z. Miao and Y. Cao. Quantifying side-chain conformational variations in protein structure. *Scientific reports*, 6(1):1–10, 2016.

[137] A. Mielke, M.A. Peletier, and D.R.M. Renger. On the relation between gradient flows and the large-deviation principle, with applications to markov chains and diffusion. *Potential Analysis*, 41(4):1293–1327, 2014.

[138] A. Mielke, D.R.M. Renger, and M.A. Peletier. A generalization of onsager's reciprocity relations to gradient flows with nonlinear mobility. *Journal of Non-Equilibrium Thermodynamics*, 41(2):141–149, 2016.

[139] M.K. Montgomery and N. Turner. Mitochondrial dysfunction and insulin resistance: an update. *Endocrine connections*, 4(1):R1–R15, 2015.

[140] J.E. Moyal. Stochastic processes and statistical physics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):150–210, 1949.

[141] J.D. Murray. *Mathematical Biology II: Spatial Models and Biomedical Applications, 3rd ed.* Springer, New York, 2003.

[142] H. Nakanishi, T. Sakaue, and J. Wakou. Hamilton-jacobi method for molecular distribution function in a chemical oscillator. *The Journal of chemical physics*, 139(21):12B602_1, 2013.

[143] G. Nicolis and R. Lefever. Comment on the kinetic potential and the maxwell construction in non-equilibrium chemical phase transitions. *Physics Letters A*, 62(7):469–471, 1977.

[144] L. Onsager. Reciprocal relations in irreversible processes. i. *Physical review*, 37(4):405, 1931.

[145] Y. Oono. Large deviation and statistical physics. *Progress of Theoretical Physics Supplement*, 99:165–205, 1989.

[146] A. Pastor and V. Hernández. Differential periodic riccati equations: existence and uniqueness of nonnegative definite solutions. *Mathematics of Control, Signals and Systems*, 6(4):341–362, 1993.

[147] R.K. Pathria and P.D. Beale. *Statistical mechanics*. Academic Press, 2011.

[148] W. Pauli. *Pauli Lectures on Physics: Thermodynamics and the Kinetic Theory of Gas*. The MIT Press M.A., 1973.

[149] M.A. Peletier, R.A. van Santen, and E. Steur. *Complexity Science: An Introduction*. World Scientific, 2019.

[150] C.H. Pence. “describing our whole experience”: The statistical philosophies of wfr weldon and karl pearson. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 42(4):475–485, 2011.

[151] L. Perko. *Differential equations and dynamical systems, 3rd ed.*, volume 7. Springer, New York, 2001, 2001.

[152] G.D.J. Phillies. *Elementary lectures in statistical mechanics*. Springer, New York, 2000.

[153] M.S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden-Day series in time series analysis. Holden-Day, 1964.

[154] S. Pressé, K. Ghosh, J. Lee, and K.A. Dill. Principles of maximum entropy and maximum caliber in statistical physics. *Reviews of Modern Physics*, 85(3):1115, 2013.

[155] H. Qian. The mathematical theory of molecular motor movement and chemomechanical energy transduction. *Journal of Mathematical Chemistry*, 27(3):219–234, 2000.

[156] H. Qian. Mathematical formalism for isothermal linear irreversibility. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 457(2011):1645–1655, 2001.

[157] H. Qian. Mesoscopic nonequilibrium thermodynamics of single macromolecules and dynamic entropy-energy compensation. *Physical Review E*, 65(1):016102, 2001.

[158] H. Qian. Relative entropy: Free energy associated with equilibrium fluctuations and nonequilibrium deviations. *Physical Review E*, 63(4):042103, 2001.

[159] H. Qian. A stochastic analysis of a brownian ratchet model for actin-based motility. *Mech. Chem. Biosyst.*, 1(4):267–278, 2004.

[160] H. Qian. Nonlinear stochastic dynamics of mesoscopic homogeneous biochemical reaction systems—an analytical theory. *Nonlinearity*, 24(6):R19, 2011.

[161] H. Qian. A decomposition of irreversible diffusion processes without detailed balance. *Journal of Mathematical Physics*, 54(5):053302, 2013.

[162] H. Qian. The zeroth law of thermodynamics and volume-preserving conservative system in equilibrium with stochastic damping. *Physics Letters A*, 378(7-8):609–616, 2014.

[163] H. Qian. Thermodynamics of the general diffusion process: Equilibrium supercurrent and nonequilibrium driven circulation with dissipation. *The European Physical Journal Special Topics*, 224(5):781–799, 2015.

[164] H. Qian. Stochastic population kinetics and its underlying mathematicothermodynamics. In *The Dynamics of Biological Systems*, pages 149–188. Springer, New York, 2019.

[165] H. Qian, P. Ao, Y. Tu, and J. Wang. A framework towards understanding mesoscopic phenomena: Emergent unpredictability, symmetry breaking and dynamics across scales. *Chemical Physics Letters*, 665:153–161, 2016.

[166] H. Qian and D.A. Beard. Metabolic futile cycles and their functions: a systems analysis of energy and control. *IEE Proceedings-Systems Biology*, 153(4):192–200, 2006.

[167] H. Qian and L.M. Bishop. The chemical master equation approach to nonequilibrium steady-state of open biochemical systems: Linear single-molecule enzyme kinetics and nonlinear biochemical reaction networks. *International journal of molecular sciences*, 11(9):3472–3500, 2010.

[168] H. Qian and Y.-C. Cheng. Counting single cells and computing their heterogeneity: from phenotypic frequencies to mean value of a quantitative biomarker. *Quantitative Biology*, 8(2):172–176, 2020.

[169] H. Qian, Y.-C. Cheng, and L.F. Thompson. Ternary representation of stochastic change and the origin of entropy and its fluctuations. *arXiv preprint arXiv:1902.09536*, 2019.

[170] H. Qian, Y.-C. Cheng, and Y.-J. Yang. Kinematic basis of emergent energetics of complex dynamics. *EPL (Europhysics Letters)*, 131(5):50002, 2020.

[171] H. Qian, Y.C. Cheng, and Y.J. Yang. Kinematic basis of emergent energetics of complex dynamics. *arXiv:1704.01828 [physics.gen-ph]*, 2020.

[172] H. Qian and H. Ge. Mesoscopic biochemical basis of isogenetic inheritance and canalization: Stochasticity, nonlinearity, and emergent landscape. *MCB: Mol. Cell. Biomech.*, 9(1):1–30, 2012.

[173] H. Qian, M. Qian, and X. Tang. Thermodynamics of the general diffusion process: time-reversibility and entropy production. *Journal of statistical physics*, 107(5):1129–1141, 2002.

[174] M. Qian and F. Zhang. Entropy production rate of the coupled diffusion process. *Journal of Theoretical Probability*, 24(3):729–745, 2011.

[175] J.D. Ramshaw. Remarks on entropy and irreversibility in non-hamiltonian systems. *Physics Letters A*, 116(3):110–114, 1986.

[176] J. Ross. *Thermodynamics and Fluctuations far from Equilibrium*. Springer, Berlin, Heidelberg, 2008.

[177] I.N. Sanov. On the probability of large deviations of random variables. Technical report, North Carolina State University. Dept. of Statistics, 1958.

[178] J.A. Schellman. Thermodynamics, molecules and the gibbs conference. *Biophysical chemistry*, 64(1-3):7–13, 1997.

[179] U. Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on progress in physics*, 75(12):126001, 2012.

[180] C.E. Shannon. The mathematical theory of communications. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[181] K. Sharp and F. Matschinsky. Translation of Ludwig Boltzmann's paper 'On the relationship between the second fundamental theorem of the mechanical theory of heat and probability calculations regarding the conditions for thermal equilibrium'. *Entropy*, 17(4):1971, 2015.

[182] G.R. Shorack. *Probability for Statisticians*. Springer Texts in Statistics. Springer International Publishing, 2017.

[183] J.E. Shore and R.W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on information theory*, 26(1):26–37, 1980.

[184] V.N. Smelyanskiy, M.I. Dykman, and R.S. Maier. Topological features of large fluctuations to the interior of a limit cycle. *Physical Review E*, 55(3):2369, 1997.

[185] E. Smith. Large-deviation principles, stochastic effective actions, path entropies, and the structure and meaning of thermodynamic descriptions. *Reports on Progress in Physics*, 74(4):046601, 2011.

[186] S.C. Stearns and R.F. Hoekstra. *Evolution, an introduction*. Oxford University Press, 2000.

[187] W.H. Steeb. Generalized liouville equation, entropy, and dynamic systems containing limit cycles. *Physica A: Statistical Mechanics and its Applications*, 95(1):181–190, 1979.

[188] D.W. Stroock and O. Zeitouni. Microcanonical distribution, Gibbs states, and the equivalence of ensembles. In Harry Kesten and Rick Durrett, editors, *Random walks, Brownian motion, and Interacting Particle Systems: A Festschrift in Honor of Frank Spitzer*, volume 28 of *Progress in Probability*, pages 399–424. Springer Science & Business Media, New York, 2012.

[189] P. Talkner and P. Hänggi. Colloquium: Statistical mechanics and thermodynamics at strong coupling: Quantum and classical. *Reviews of Modern Physics*, 92(4):041002, 2020.

[190] T. Tao. *Topics in Random Matrix Theory*. Graduate studies in mathematics. American Mathematical Society, 2012.

[191] H. Tasaki. On the local equivalence between the canonical and the microcanonical ensembles for quantum spin systems. *Journal of Statistical Physics*, 172(4):905–926, 2018.

[192] Y.L. Tong. *The Multivariate Normal Distribution*. Springer Science & Business Media, 2012.

[193] H. Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1-3):1–69, 2009.

[194] H. Touchette. Ensemble equivalence for general many-body systems. *EPL (Europhysics Letters)*, 96(5):50010, 2011.

[195] H. Touchette. Equivalence and nonequivalence of ensembles: Thermodynamic, macrostate, and measure levels. *Journal of Statistical Physics*, 159(5):987–1016, 2015.

[196] J. Van Campenhout and T. Cover. Maximum entropy and conditional probability. *IEEE Transactions on Information Theory*, 27(4):483–489, 1981.

[197] C. Van den Broeck and M. Esposito. Three faces of the second law. ii. fokker-planck formulation. *Physical Review E*, 82(1):011144, 2010.

[198] C. Van den Broeck and M. Esposito. Ensemble and trajectory thermodynamics: A brief introduction. *Physica A: Statistical Mechanics and its Applications*, 418:6–16, 2015.

[199] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, 1992.

[200] W. Vance and J. Ross. Fluctuations near limit cycles in chemical reaction systems. *The Journal of chemical physics*, 105(2):479–487, 1996.

[201] O.A. Vasicek. A conditional law of large numbers. *The Annals of Probability*, pages 142–147, 1980.

[202] M. Vellela and H. Qian. Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the schlögl model revisited. *Journal of The Royal Society Interface*, 6(39):925–940, 2009.

[203] J. Venn. *The Logic of Chance, 3rd*. Macmillan & Co, London, 1888.

[204] A.D. Ventsel and M.I. Freidlin. On small random perturbations of dynamical systems. *Russian Mathematical Surveys*, 25(1):1–55, 1970.

[205] L. von Bertalanffy. The theory of open systems in physics and biology. *Science*, 111(2872):23–29, 1950.

[206] D. Wallace. Naturalness and emergence, February 2019.

[207] J. Wang. Landscape and flux theory of non-equilibrium dynamical systems with application to biology. *Advances in Physics*, 64(1):1–137, 2015.

[208] J. Wang, L. Xu, and E. Wang. Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. *Proceedings of the National Academy of Sciences*, 105(34):12271–12276, 2008.

[209] J. Wang, L. Xu, and E. Wang. Robustness, dissipations and coherence of the oscillation of circadian clock: potential landscape and flux perspectives. *PMC biophysics*, 1(1):7, 2008.

[210] J. Wang, K. Zhang, and E. Wang. Kinetic paths, time scale, and underlying landscapes: A path integral framework to study global natures of nonequilibrium systems and networks. *The Journal of chemical physics*, 133(12):09B613, 2010.

[211] E. Widmaier, H. Raff, and K. Strang. *Human physiology: the mechanisms of body function, 13th Edition*. McGraw-Hill, 2014.

[212] Z. Xiang, P.J. Steinbach, M.P. Jacobson, R.A. Friesner, and B. Honig. Prediction of side-chain conformations on protein surfaces. *Proteins: Structure, Function, and Bioinformatics*, 66(4):814–823, 2007.

[213] L. Xu, F. Zhang, E. Wang, and J. Wang. The potential and flux landscape, lyapunov function and non-equilibrium thermodynamics for dynamic systems and networks with an application to signal-induced ca2+ oscillation. *Nonlinearity*, 26(2):R69, 2012.

[214] L. Xu, F. Zhang, K. Zhang, E. Wang, and J. Wang. The potential and flux landscape theory of ecology. *PLoS One*, 9(1):e86746, 2014.

[215] F.X.-F. Ye and H. Qian. Stochastic dynamics ii: finite random dynamical systems, linear representation, and entropy production. *Discrete and Continuous Dynamical Systems Series B*, 24(8):4341–4366, 2019.

[216] L.-S. Young. What are srb measures, and which dynamical systems have them? *Journal of statistical physics*, 108(5):733–754, 2002.

[217] S.L. Zabell. Rates of convergence for conditional expectations. *The Annals of Probability*, pages 928–941, 1980.

[218] F. Zhang, L. Xu, and J. Wang. The extinction differential induced virulence macroevolution. *Chemical Physics Letters*, 599:38–43, 2014.

[219] F. Zhang, L. Xu, K. Zhang, E. Wang, and J. Wang. The potential and flux landscape theory of evolution. *The Journal of chemical physics*, 137(6):065102, 2012.

[220] X.-J. Zhang, H. Qian, and M. Qian. Stochastic theory of nonequilibrium steady states and its applications. part i. *Physics Reports*, 510(1-2):1–86, 2012.

[221] B. Zhou and G.R. Duan. Periodic lyapunov equation based approaches to the stabilization of continuous-time periodic linear systems. *IEEE Transactions on Automatic Control*, 57(8):2139–2146, 2011.

[222] P. Zhou and T. Li. Construction of the landscape for multi-stable systems: Potential landscape, quasi-potential, a-type integral and beyond. *The Journal of Chemical Physics*, 144(9):094109, 2016.