The Annals of Probability
2020, Vol. 48, No. 5, 2525–2564
https://doi.org/10.1214/20-AOP1430

LARGE DEVIATIONS AND LOCALIZATION OF THE MICROCANONICAL ENSEMBLES GIVEN BY MULTIPLE CONSTRAINTS

BY KYEONGSIK NAM

Department of Mathematics, University of California, Berkeley, ksnam@math.berkeley.edu

We develop a unified theory to analyze the microcanonical ensembles with several constraints given by unbounded observables. Several interesting phenomena that do not occur in the single constraint case can happen under the multiple constraints case. We systematically analyze the detailed structures of such microcanonical ensembles in two orthogonal directions using the theory of large deviations. First of all, we establish the equivalence of ensembles result, which exhibits an interesting phase transition phenomenon. Secondly, we study the localization and delocalization phenomena by obtaining large deviation results for the joint law of empirical distributions and the maximum component. Some concrete examples for which the theory applies will be given as well.

1. Introduction.

1.1. *Motivation.* There are several notions of statistical ensembles describing the mechanical system. For instance, a canonical ensemble represents the possible states in the equilibrium with a heat reservoir at a fixed temperature, whereas a microcanonical ensemble represents the states having a specified total energy. The Gibbs' principle, which is also called the principle of *equivalence of ensembles*, states that in the infinite volume limit, the microcanonical ensemble converges to the canonical ensemble with a certain temperature. A theory of large deviations has provided an elegant way to describe the equivalence of ensembles results. We refer to [5, 9, 14] for a monograph about Gibbs measures and the statistical mechanics.

The theory of microcanonical ensembles with a single constraint has been well established. For instance, consider the microcanonical ensemble given by a uniform distribution on the set

$$\left\{\left|\frac{\phi(x_1)+\cdots+\phi(x_n)}{n}-c\right|\le\delta\right\},\quad \delta>0 \text{ small,} \tag{1}$$

where a random vector distributed as a uniform distribution on (1) is denoted by $X=(X_1,\dots,X_n)$ (a random vector notation X will be used throughout the Introduction). The principle of equivalence of ensembles asserts that the law of X_1 converges weakly to the probability distribution λ^* maximizing a differential entropy $h(\mu)$ over the constraint $\int \phi\, d\mu = c$, as $n\to\infty$ followed by $\delta\to 0$ (see Proposition 5.1 for details).

Beyond the single constraint of type (1), it is natural and crucial to consider the microcanonical ensembles given by *several* constraints:

$$\bigcap_{i=1}^{2}\left\{\left|\frac{\phi_i(x_1)+\cdots+\phi_i(x_n)}{n}-a_i\right|\le\delta\right\},\quad \delta>0 \text{ small.} \tag{2}$$

Received February 2019; revised February 2020.
MSC2020 subject classifications. 60F10, 82B05, 82B26.
Key words and phrases. Large deviation principle, equivalence of ensembles, microcanonical ensembles, phase transition, localization.

Here, the configuration space $\Omega := (0,\infty)^n$ is equipped with the reference measure $\mathbb{P} := \lambda^{\otimes n}$ for some probability measure λ on $(0,\infty)$. The microcanonical ensemble is a conditional distribution of $\mathbb{P}$ on the set (2). The classical Gibbs principle for the microcanonical ensemble with a single constraint naturally applies if ϕ_i's are *bounded* and continuous: the law of X_1 converges weakly to the distribution $d\lambda^* = \frac{1}{Z}e^{\alpha\phi_1+\beta\phi_2}\,d\lambda$ for some α,β satisfying $\int \phi_i\,d\lambda^* = a_i$, $i=1,2$, as $n\to\infty$ followed by $\delta\to 0$.

However, entirely new and interesting phenomena happen if ϕ_i's are *not* bounded. For instance, for the configuration space $(0,\infty)^n$, consider the uniform distribution on the set

$$\left\{\frac{\phi_1(x_1)+\cdots+\phi_1(x_n)}{n}=1\right\}\cap\left\{\frac{\phi_2(x_1)+\cdots+\phi_2(x_n)}{n}=b\right\} \tag{3}$$

with unbounded functions $\phi_i(x)=x^i$ for $i=1,2$. Chatterjee [2] proved that when $1\le b\le 2$, the law of X_1 converges weakly to the $G_{1,b}$-distribution as $n\to\infty$, where $G_{1,b}$ is a probability distribution on $(0,\infty)$ of the form $\frac{1}{Z}e^{rx+sx^2}\,dx$ satisfying

$$\int x\,dG_{1,b}=1,\qquad \int x^2\,dG_{1,b}=b.$$

On the other hand, a genuinely new phenomenon appears when $b>2$. In fact, the law of X_1 converges weakly to exp(1) distribution as $n\to\infty$, whatever the value of $b>2$ is. This explains that one of the constraints becomes irrelevant to the thermodynamic behavior of (3). Another striking fact is that the expectation of ϕ_2 under the limiting distribution, which is equal to 2, is strictly less than b. In other words, some localized site $1\le i\le n$ possesses a strictly positive l^2-mass $\frac{x_i^2}{n}$ and a negligible l^1-mass $\frac{x_i}{n}$ due to the existence of a discrepancy $b-2$ corresponding to the second constraint. This example shows that microcanonical ensembles given by several constraints (2) with unbounded ϕ_i's behave qualitatively differently from the ensembles with bounded ϕ_i's.

The microcanonical ensembles with several constraints, given by unbounded observables such as (3), are of great importance in the statistical mechanics due to their wide applications to the partial differential equations. For instance, consider the (focusing) nonlinear Schrödinger equation (NLS)

$$u_t = -\Delta u - |u|^{p-1}u \tag{4}$$

(see [17] for a monograph of dispersive equations). Since the Hamiltonian $H(u) = \int \frac{1}{2}|\nabla u|^2 - \frac{1}{p+1}|u|^{p+1}$ and the mass $M(u)=\int |u|^2$ are conserved under the NLS (4), a natural invariant measure for NLS is a uniform distribution on the set

$$\bigl|M(u)-m\bigr|\le\delta,\qquad \bigl|H(u)-E\bigr|\le\delta,\quad \delta>0\text{ small}. \tag{5}$$

This microcanonical ensemble can be made rigorous by discretizing the underlying space $\mathbb{R}^d$. Chatterjee [1] showed that in a suitable thermodynamic limit, a typical function in the microcanonical ensemble (5) approximates the ground state soliton. As a consequence, a statistical version of the soliton resolution conjecture was established. This example illustrates that thermodynamic properties of a general class of microcanonical ensembles can lead to describe a probabilistic behavior of the corresponding PDEs.

However, despite its importance, no systematic methods to analyze such microcanonical ensembles have been developed yet. To the author's best knowledge, the only concrete examples of such microcanonical ensembles that have been studied so far are (3) with $\phi_i(x)=x^i$ (see [2]) and (5) (see [1]). However, methods used there are ad hoc, and finer structures of the microcanonical ensembles are far from being well understood. In this paper, we first develop a unified framework describing thermodynamic behaviors of general microcanonical ensembles with multiple constraints given by unbounded functions. Remarkably, it turns out that such microcanonical ensembles behave significantly differently from the single constraint case or the multiple constraints given by bounded observables.

1.2. *Previous works and main contributions of our work.* The equivalence result between the microcanonical ensemble given by a single constraint and the grand canonical ensemble is quite classical and has been studied extensively. We refer to [4, 13] for the case when the Hamiltonian describing a constraint is given by a bounded interacting potential, and [6–8] for the case including a possibly unbounded interacting potential. The similar equivalence of ensembles result holds for the microcanonical ensemble with multiple constraints of type

$$\bigcap_{i=1}^{k}\left\{\left|\frac{\phi_i(X_1)+\cdots+\phi_i(X_n)}{n}-a_i\right|\le\delta\right\},\quad k\ge 2,\delta>0\text{ small},\tag{6}$$

provided that ϕ_i's are bounded functions.

However, as explained before, an entirely new phenomenon occurs if ϕ_i's are not bounded functions: as mentioned before, some of multiple constraints in (6) may become extraneous in a thermodynamic limit. The first main contribution of our work is to develop a unified framework about the equivalence of ensembles result for general microcanonical ensembles with several constraints given by unbounded observables. The framework developed in this paper is new in the statistical mechanics literature, and robust which clearly explains a mysterious thermodynamic behavior of such microcanonical ensembles. We build a framework using a large deviation theory, and precisely characterize a thermodynamic limit of such microcanonical ensembles.

Another entirely new phenomenon of the microcanonical ensembles given by several constraints with unbounded observables is a *localization* phenomenon, which complements the equivalence of ensembles result (see Section 2 for the explanations). Under the single constraint (1) with an unbounded function ϕ, it is not hard to see that localization does not happen (see Proposition 5.2 for a precise statement and a proof). On the other hand, for the microcanonical ensemble given by multiple constraints (6) with unbounded functions ϕ_i's, as explained in the example (3), a strictly positive mass can be concentrated on some sites (see Theorem 2.8 for details). The second contribution of our paper is a systematic study on the localization and delocalization phenomena of the general microcanonical ensembles. In particular, we derive a large deviation principle for the joint law of empirical distributions and the maximum component. This type of result is new in the statistical mechanics literature, and reveals a genuinely new and detailed structure of such microcanonical ensembles.

1.3. *Organization of the paper.* The paper is organized as follows. In Section 2, we first introduce a precise model of microcanonical ensembles with several constraints given by unbounded observables, and then state main theorems and explain their interpretations.

The key principle behind a new behavior of such microcanonical ensembles is a large deviation result for the joint law of empirical distributions and several empirical means (Theorem 3.4). Using this result combined with the Gibbs conditioning principle, we precisely characterize a thermodynamic limit of microcanonical ensembles. Details are elaborated in Section 3.

In Section 4, we describe the localization and delocalization phenomena of microcanonical ensembles. The key result is a large deviation result for the joint law of empirical distributions and the maximum component (Theorem 4.1), which illustrates a detailed thermodynamic behavior of the microcanonical ensembles. Finally, in Section 5, some concrete examples of the microcanonical distributions for which the theory applies will be covered.

1.4. *Notations.* Throughout the paper, for a Polish space $\mathcal{S}$, let us denote $\mathcal{B}$ by the Borel σ-field on $\mathcal{S}$ and $C_b(\mathcal{S})$ by the set of bounded continuous functions on $\mathcal{S}$. Let us define

$\mathcal{M}(\mathcal{S})$ as the set of finite regular Borel measures on $\mathcal{S}$, and $\mathcal{M}_1(\mathcal{S})$ as the subspace of probability measures. Given the set of bounded continuous functions $\{g_k\}$ that determine the weak convergence on $\mathcal{M}_1(\mathcal{S})$, we define a metric d on $\mathcal{M}_1(\mathcal{S})$ by

$$d(\mu,\nu) := \sum_{k=1}^{\infty} \frac{1}{2^k \|g_k\|_\infty} \left[\int_{\mathcal{S}} g_k \, d\mu - \int_{\mathcal{S}} g_k \, d\nu \right] \tag{7}$$

for two probability measures μ, ν. Note that the weak topology on $\mathcal{M}_1(\mathcal{S})$ coincides with the topology given by a metric d. Throughout this paper, we assume that $\mathcal{M}_1(\mathcal{S})$ is equipped with the weak topology. For $\mu, \nu \in \mathcal{M}_1(\mathcal{S})$, $H(\mu|\nu)$ denotes the *relative entropy* between μ and ν. Finally, ∂f denotes a subdifferential of the function f, and the integral $\int_0^\infty f$ is simplified to $\int f$.

2. Main results. Consider the configuration space $\Omega = (0,\infty)^{\mathbb{N}}$, and let us denote $X_i : \Omega \to (0,\infty)$ by the projection onto the ith coordinate. Assume that the functions $\phi_1, \ldots, \phi_k$ ($k \geq 2$) satisfying the following Assumption 1 are given.

ASSUMPTION 1. Functions $\phi_1, \ldots, \phi_k$ ($k \geq 2$) satisfy:

(i) For each $1 \leq i \leq k$, $\phi_i : (0,\infty) \to (0,\infty)$ is C^1, increasing, and $\lim_{x\to\infty} \phi_i(x) = \infty$.
(ii) For each $1 \leq i \leq k$ and any $c > 0$, $\int_0^\infty e^{-c\phi_i} \, dx < \infty$.
(iii) There exists $\kappa > 1$ such that $\phi_i^\kappa < \phi_{i+1}$ for each $1 \leq i \leq k-1$.
(iv) There exists $C, M > 0$ such that $x > C \Rightarrow \frac{1}{C}\phi_i(x)^{-M} < \phi_i'(x) < C\phi_i(x)^M$ for each i.

Conditions (i) and (ii) imply that ϕ_i's are unbounded and grows not slowly at infinity. Condition (iii) means that for each index i, ϕ_{i+1} grows faster than ϕ_i at infinity. A technical assumption (iv) will be used to prove Lemma A.1 later. It is not hard to see that a large class of functions ϕ_i's satisfy the Assumption 1. For instance, a large class of polynomials with strictly increasing degrees, which is of our main interest due to its wide applications in geometry and PDEs as explained in the Introduction, satisfy the Assumption 1. In particular, the constraint that Chatterjee considered in [2] corresponds to the case $\phi_1(x) = x$ and $\phi_2(x) = x^2$.

For each $1 \leq i \leq k$, define the empirical means

$$S_n^i := \frac{\phi_i(X_1) + \cdots + \phi_i(X_n)}{n},$$

and then consider the following constraints for each $\delta > 0$:

$$C_n^\delta := \bigcap_{i=1}^{k} \{|S_n^i - a_i| \leq \delta\}.$$

We are interested in the infinite volume behavior of the uniform distribution on the constraint C_n^δ as the gap δ converges to zero. Since the Lebesgue measure is not a probability measure, we define a reference measure $\mathbb{P}$ to be $\mathbb{P} := \lambda^{\otimes\mathbb{N}}$ on $\Omega = (0,\infty)^{\mathbb{N}}$, where λ is a probability measure on $(0,\infty)$ defined by

$$\lambda = \frac{1}{Z} e^{-\phi_1} \, dx \tag{8}$$

(Z is a normalizing constant). The motivation to choose such reference measure is that it is a probability measure and once conditioned on the constraint C_n^δ, it behaves like the uniform distribution as $\delta \to 0$. In fact, the conditional distribution of any reference measure $(\frac{1}{Z} e^{p_1\phi_1 + \cdots + p_k\phi_k} \, dx)^{\otimes\mathbb{N}}$ on the constraint C_n^δ approximates the uniform distribution in a certain sense as $\delta \to 0$. We refer to Remark 2.7 for the detailed explanations.

Now, let us consider the following microcanonical distribution:

$$\mathbb{P}\big((X_1,\dots,X_n)\in\cdot|C_n^\delta\big). \tag{9}$$

We develop a unifying method to systematically analyze the detailed behaviors of (9) as $n\to\infty$ followed by $\delta\to 0$.

Note that for certain values of $(a_1,\dots,a_k)$, the conditional distribution (9) may not be well defined since the constraint C_n^δ may be an empty set for small $\delta>0$. In order to avoid this problem, we define the *admissible set* in the following way: let us denote $\mathcal{A}_1\subset(0,\infty)^{(k-1)}$ by

$$\mathcal{A}_1 := \operatorname{int}\bigg\{(v_1,\dots,v_{k-1})\in(0,\infty)^{(k-1)}:\exists\mu\in\mathcal{M}_1(\mathbb{R}^+)\text{ such that}$$
$$h(\mu)\neq-\infty,\int\phi_1\,d\mu=v_1,\dots,\int\phi_{k-1}\,d\mu=v_{k-1},\int\phi_k\,d\mu<\infty\bigg\}.$$

Here, $h(\mu)$ is the *differential entropy* of the probability measure $\mu\in\mathcal{M}_1(\mathbb{R}^+)$, defined by

$$h(\mu):=\begin{cases}-\displaystyle\int\frac{d\mu}{dx}\log\left(\frac{d\mu}{dx}\right)dx & \mu\ll dx,\\ -\infty & \text{otherwise.}\end{cases}$$

For each $(v_1,\dots,v_{k-1})\in\mathcal{A}_1$, define

$$g_1(v_1,\dots,v_{k-1}) := \inf_{\mu\in\mathcal{M}_1(\mathbb{R}^+)}\bigg\{\int\phi_k\,d\mu:h(\mu)\neq-\infty,\int\phi_1\,d\mu=v_1,\dots,\int\phi_{k-1}\,d\mu=v_{k-1}\bigg\}.$$

Finally, the admissible set $\mathcal{A}$ is defined by

$$\mathcal{A}:=\big\{(v_1,\dots,v_{k-1},v_k):(v_1,\dots,v_{k-1})\in\mathcal{A}_1, v_k>g_1(v_1,\dots,v_{k-1})\big\}.$$

Also, we assume that a map $g_1:\mathcal{A}_1\to\mathbb{R}$ is continuous, which implies that $\mathcal{A}$ is an open set. Throughout this paper, we only consider the case $(a_1,\dots,a_k)\in\mathcal{A}$ so that the constraint C_n^δ is a nonempty set, and thus the microcanonical distribution is well defined (see Remark 3.8 for the explanations).

We first characterize the law to which the finite marginal distribution $\mathbb{P}((X_1,\dots,X_j)\in\cdot|C_n^\delta)$ weakly converges as $n\to\infty$ followed by $\delta\to 0$.

THEOREM 2.1. *Let λ^* be the (unique) maximizer of the differential entropy $h(\cdot)$ over the set*

$$\bigg\{\mu\in\mathcal{M}_1(\mathbb{R}^+):\int\phi_1\,d\mu=a_1,\dots,\int\phi_{k-1}\,d\mu=a_{k-1},\int\phi_k\,d\mu\le a_k\bigg\}. \tag{10}$$

Then, for any fixed positive integer j,

$$\lim_{\delta\to0}\lim_{n\to\infty}\mathbb{P}\big((X_1,\dots,X_j)\in\cdot|C_n^\delta\big)=(\lambda^*)^{\otimes j}. \tag{11}$$

REMARK 2.2. When each function ϕ_i is bounded and continuous, as a simple application of the maximum entropy principle, one can deduce that the limiting law λ^* in (11) is a (unique) maximizer of the differential entropy $h(\cdot)$ over the set

$$\bigg\{\mu\in\mathcal{M}_1(\mathbb{R}^+):\int\phi_i\,d\mu=a_i, 1\le i\le k\bigg\}. \tag{12}$$

In fact, according to the maximum entropy principle, the limiting distribution λ^* in (11) is a (unique) minimizer of the relative entropy $H(\cdot|\lambda)$ over the set (12). Thus, using the identity, for $\mu \ll dx$,

$$\begin{aligned} H(\mu|\lambda) &= \int \log\left(\frac{d\mu}{d\lambda}\right) d\mu \\ &= \int \log\left(\frac{d\mu}{dx}\right) d\mu + \int \log\left(\frac{dx}{d\lambda}\right) d\mu \\ &= -h(\mu) + a_1 + C, \end{aligned}$$

it follows that λ^* is a (unique) maximizer of the differential entropy $h(\cdot)$ over the set (12).

On the other hand, when the macroscopic observables ϕ_i's are unbounded, the classical maximum entropy principle is not applicable since the map $\mu \mapsto \int \phi_i \, d\mu$ may not be continuous. Theorem 2.1 claims that the last condition $\int \phi_k \, d\mu = a_k$ in the set (12) is enlarged to the condition $\int \phi_k \, d\mu \le a_k$. This implies that for certain values of $a_1, \dots, a_{k-1}$, the last constraint $|S_n^k - a_k| \le \delta$ may be irrelevant to the limiting law of the finite marginal distribution of (9). In other words, unlike the case when ϕ_i's are bounded continuous, the limiting law λ^* may not satisfy $\int \phi_k \, d\lambda^* = a_k$. This phenomenon is precisely described in Theorem 2.4, which is about the equivalence of ensembles result.

It turns out that as in Remark 2.2, the structure of a set (10) in Theorem 2.1 is also different from the case when the microcanonical distribution is given by a single constraint. In fact, in the case of single constraint (1) under the reference measure $(\frac{1}{Z} e^{-\phi} \, dx)^{\otimes \mathbb{N}}$ (assume that an unbounded function ϕ satisfies the conditions (i) and (ii) in Assumption 1), λ^* in Theorem 2.1 is given by

$$\lambda^* = \underset{\mu \in \mathcal{M}_1(\mathbb{R}^+)}{\arg\max} \left\{ h(\mu) : \int \phi \, d\mu \le c \right\} = \underset{\mu \in \mathcal{M}_1(\mathbb{R}^+)}{\arg\max} \left\{ h(\mu) : \int \phi \, d\mu = c \right\}$$

(see Section 5.1 and the identity (89)). In other words, even when ϕ is unbounded, the limiting distribution λ^* satisfies $\int \phi \, d\lambda^* = c$. We refer to [7, 8] for the similar equivalence of ensembles result for more general Hamiltonian with superstable interactions.

On the other hand, as mentioned before, in the case of multiple constraints with unbounded ϕ_i's satisfying Assumption 1, the expectation of ϕ_k under the limiting distribution λ^* may not be equal to a_k. Also, the expectation of ϕ_i's $(1 \le i \le k-1)$ under the limiting distribution λ^* is always equal to a_i. This is because the unbounded function ϕ_k controls other functions, and the main reason behind this phenomenon is illustrated in Theorem 3.4.

Now, let us precisely characterize a unique maximizer of the differential entropy $h(\cdot)$ over the set (10). In order to accomplish this, we need the following definition.

DEFINITION 2.3. Define the logarithmic moment generating function:

$$H(p_1, \dots, p_k) := \log \int e^{p_1\phi_1 + \dots + p_k\phi_k} \, d\lambda. \tag{13}$$

Let us denote π_1 and π_2 by the projections $\pi_1(v_1, \dots, v_{k-1}, v_k) = (v_1, \dots, v_{k-1})$ and $\pi_2(v_1, \dots, v_{k-1}, v_k) = v_k$. Then, define $\mathcal{S}_1 \subset \mathcal{A}_1$ by a collection of $(v_1, \dots, v_{k-1})$'s such that there exist $p_1, \dots, p_{k-1}$ satisfying

$$(v_1, \dots, v_{k-1}) \in \pi_1(\partial H(p_1, \dots, p_{k-1}, 0)). \tag{14}$$

For $(v_1, \dots, v_{k-1}) \in \mathcal{S}_1$, choose a unique $(p_1, \dots, p_{k-1})$ satisfying (14) (see Remark 3.11 for the explanations), and then define a function $g_2 : \mathcal{S}_1 \to \mathbb{R}$ by

$$g_2(v_1, \dots, v_{k-1}) := \inf\{\pi_2(\partial H(p_1, \dots, p_{k-1}, 0))\}. \tag{15}$$

Finally, define $\mathcal{S}_2 := \mathcal{A}_1 \cap \mathcal{S}_1^c$.

Now, one can precisely characterize the distribution λ^* in Theorem 2.1 using the notions in Definition 2.3. It exhibits an interesting phase transition phenomenon.

THEOREM 2.4. *Fix any positive integer j. Then,*

$$\lim_{\delta\to 0}\lim_{n\to\infty}\mathbb{P}\big((X_1,\dots,X_j)\in\cdot|C_n^\delta\big)=\big(\lambda^*\big)^{\otimes j},$$

where λ^* is characterized as follows: either in the case of

(i) $(a_1,\dots,a_{k-1})\in\mathcal{S}_2$ *or*
(ii) $(a_1,\dots,a_{k-1})\in\mathcal{S}_1$ *and* $a_k<g_2(a_1,\dots,a_{k-1})$,

$$\lambda^*=\frac{1}{Z}e^{p_1\phi_1+\dots+p_{k-1}\phi_{k-1}+p_k\phi_k}\,dx$$

for $p_1,\dots,p_{k-1},p_k$ satisfying $p_k<0$ and $\int\phi_i\,d\lambda^*=a_i$ for $1\le i\le k$.

On the other hand, when $(a_1,\dots,a_{k-1})\in\mathcal{S}_1$ and $a_k\ge g_2(a_1,\dots,a_{k-1})$,

$$\lambda^*=\frac{1}{Z}e^{p_1\phi_1+\dots+p_{k-1}\phi_{k-1}}\,dx$$

for $p_1,\dots,p_{k-1}$ satisfying $\int\phi_i\,d\lambda^*=a_i$ for $1\le i\le k-1$.

According to the Gibbs principle, if each ϕ_i is bounded continuous, then the limiting law is of the form $\lambda^*=\frac{1}{Z}e^{p_1\phi_1+\dots+p_k\phi_k}\,dx$ satisfying $\int\phi_i\,d\lambda^*=a_i$ for all $1\le i\le k$. Also, when the microcanonical ensemble is given by a single constraint (1), even when ϕ is not bounded, one can prove a similar result (see Proposition 5.1). However, when the constraints are given by several unbounded observables satisfying Assumption 1, Theorem 2.4 demonstrates that one of the constraints may not contribute to the limiting distribution λ^*. We refer to Sections 5.2 and 5.3 for some concrete examples.

Theorem 2.4 also shows that the interesting phase transition phenomenon happens in the equivalence of ensembles viewpoint. Indeed, when $(a_1,\dots,a_{k-1})\in\mathcal{S}_1$ and $a_k\ge g_2(a_1,\dots,a_{k-1})$, the kth constraint $S_n^k=a_k$ becomes extraneous for a limit of the finite marginal distributions of the microcanonical ensembles. Since λ^* in Theorem 2.4 satisfies $\int\phi_k\,d\lambda^*=g_2(a_1,\dots,a_{k-1})$ (see Lemma 3.12), it is plausible to guess that the discrepancy $a_k-g_2(a_1,\dots,a_{k-1})$ corresponding to the kth constraint gets concentrated on some sites. We will rigorously elaborate on this point in Theorem 2.8.

On the other hand, when $(a_1,\dots,a_{k-1})\in\mathcal{S}_1$ and $a_k<g_2(a_1,\dots,a_{k-1})$, in the equivalence of ensembles viewpoint Theorem 2.4, the microcanonical distributions (9) behave in a standard way. In other words, as in the case when ϕ_i's are bounded, the limiting distribution λ^* satisfies $\int\phi_i\,d\lambda^*=a_i$ for all $1\le i\le k$. From this, we can infer that no huge amount of the quantity can be concentrated on some sites (see Theorem 2.6 for the precise statement), unlike the case $a_k\ge g_2(a_1,\dots,a_{k-1})$. Another interesting point of Theorem 2.4 is the case when $(a_1,\dots,a_{k-1})\in\mathcal{S}_2$: unlike the case $(a_1,\dots,a_{k-1})\in\mathcal{S}_1$, whatever a_k is, the limiting distribution λ^* satisfies $\int\phi_i\,d\lambda^*=a_i$ for all $1\le i\le k$. This key difference of the sets $\mathcal{S}_1$ and $\mathcal{S}_2$ follows from Lemma 3.12.

Although Theorem 2.4 provides the equivalence of ensembles result and explains the interesting phase transition phenomenon, it does not capture the localization phenomenon. In order to illustrate this, assume for a moment that in a thermodynamic limit, a huge amount of the quantity gets concentrated on a single site. It is obvious that the probability that this localized site is the first coordinate of the configuration space is equal to $\frac{1}{n}$. Since $\frac{1}{n}$ converges to zero as $n\to\infty$, this localization phenomenon is not reflected in the statement

$$\lim_{\delta\to 0}\lim_{n\to\infty}\mathbb{P}\big(X_1\in\cdot|C_n^\delta\big)=\lambda^*.$$

Therefore, the localization and delocalization phenomena can provide the supplementary information about the microcanonical ensembles. We study this phenomenon by obtaining a large deviation result for the maximum component. In fact, we can analyze much finer structures of (9) by establishing a large deviation result for the joint law of empirical distributions $L_n := \frac{1}{n}(\delta_{X_1} + \cdots + \delta_{X_n})$ and the maximum component $M_n := \max_{1\le i\le n} \frac{\phi_k(X_i)}{n}$ under the microcanonical distribution (9).

THEOREM 2.5. *For any Borel set A in $\mathcal{M}_1(\mathbb{R}^+) \times \mathbb{R}^+$,*

$$\begin{aligned}
&-\inf_{(\mu,z)\in A^{\circ}} J^{\max}(\mu, z)\\
&\quad\le \liminf_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big((L_n, M_n)\in A^{\circ}|C_n^{\delta}\big)\\
&\quad\le \limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big((L_n, M_n)\in \bar{A}|C_n^{\delta}\big) \le -\inf_{(\mu,z)\in\bar{A}} J^{\max}(\mu, z)
\end{aligned}$$

with the rate function $J^{\max}(\mu, z)$ given by

$$\begin{aligned}
&J^{\max}(\mu, z)\\
&\quad=\begin{cases} -h(\mu) - K(a_1,\dots,a_k) & \int \phi_i\,d\mu = a_i (1\le i\le k-1), \int \phi_k\,d\mu \le a_k - z,\\ \infty & \text{otherwise}.\end{cases}
\end{aligned}$$

Here, $K(a_1,\dots,a_k)$ is defined by

$$\begin{aligned}
&K(a_1,\dots,a_k)\\
&\quad:= \inf_{\mu\in\mathcal{M}_1(\mathbb{R}^+)}\left\{-h(\mu) : \int\phi_1\,d\mu = a_1,\dots,\int\phi_{k-1}\,d\mu = a_{k-1}, \int\phi_k\,d\mu\le a_k\right\}.
\end{aligned}$$

Theorem 2.5 provides fine structures of the microcanonical ensembles since it offers the limit behaviors of the joint law of empirical distributions and the maximum component. In particular, one can systematically analyze the localization and delocalization phenomena of the microcanonical ensembles using the large deviation result Theorem 2.5. First, one can prove the following delocalization result.

THEOREM 2.6. *Fix any $\epsilon > 0$. Then, either in the case of*

(i) *$(a_1,\dots,a_{k-1}) \in \mathcal{S}_2$ or*
(ii) *$(a_1,\dots,a_{k-1}) \in \mathcal{S}_1$ and $a_k \le g_2(a_1,\dots,a_{k-1})$,*

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(M_n\ge\epsilon|C_n^{\delta}\big) < 0. \tag{16}$$

In particular, localization does not happen in the sense that

$$\lim_{\delta\to 0}\lim_{n\to\infty}\mathbb{P}\big(M_n<\epsilon|C_n^{\delta}\big) = 1. \tag{17}$$

On the other hand, in the case of $(a_1,\dots,a_{k-1}) \in \mathcal{S}_1$ and $a_k > g_2(a_1,\dots,a_{k-1})$, we have the upper tail estimate for the maximum component:

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(M_n\ge a_k - g_2(a_1,\dots,a_{k-1})+\epsilon|C_n^{\delta}\big) < 0. \tag{18}$$

In particular, the maximum component cannot be too large in the sense that

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\big(M_n < a_k - g_2(a_1,\dots,a_{k-1})+\epsilon|C_n^{\delta}\big) = 1.$$

Theorem 2.6 claims that for certain values of $(a_1, \dots, a_k)$ (condition (i) or (ii) in Theorem 2.6), delocalization happens in the sense that (17) holds. On the other hand, when $(a_1, \dots, a_{k-1}) \in \mathcal{S}_1$ and $a_k > g_2(a_1, \dots, a_{k-1})$, as we predicted before, it is plausible to expect that the localization phenomenon happens. Since Theorem 2.6 provides the upper tail estimate (18) for the maximum component, if we have an analogous lower tail estimate:

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(M_n \le a_k - g_2(a_1,\dots,a_{k-1}) - \epsilon | C_n^\delta\big) < 0, \tag{19}$$

then we can deduce that M_n approximates to $a_k - g_2(a_1, \dots, a_{k-1})$ as $n \to \infty$ followed by $\delta \to 0$, which implies the localization phenomenon. Unfortunately, using the large deviation result Theorem 2.5, one can check that (19) is false in general: indeed,

$$\lim_{\delta\to 0}\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(M_n \le a_k - g_2(a_1,\dots,a_{k-1}) - \epsilon | C_n^\delta\big) = 0 \tag{20}$$

(see Section 4.2 for the explanations). Therefore, in order to obtain the lower tail estimate of type (19), we need to scale down the scaling factor n. Remarkably, it turns out that unlike the upper tail estimate (18) or the delocalization estimate (16), the correct scaling factor in (19) highly depends on the detailed structures of the functions ϕ_i's:

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{g(n)}\log\mathbb{P}\big(M_n \le a_k - g_2(a_1,\dots,a_{k-1}) - \epsilon | C_n^\delta\big) < 0, \tag{21}$$

for some function g heavily relying on ϕ_i's. Also, since the scaling factor $g(n)$ grows slowly than n, unlike the estimate (16) or (18), the left-hand side of the lower tail estimate (21) is sensitive to the particular choice of the reference measure of the form

$$\left(\frac{1}{Z}e^{p_1\phi_1+\dots+p_k\phi_k}\,dx\right)^{\otimes\mathbb{N}}$$

due to the following Remark 2.7.

REMARK 2.7. We have developed theories under the particular reference measure $\mathbb{P} = \lambda^{\otimes\mathbb{N}}$ with λ given by (8) since it is a probability measure and once conditioned on the constraint C_n^δ, it behaves like the uniform distribution, which is of our main interest. In order to explain this rigorously, let us consider the probability measure ν on $(0, \infty)$ given by

$$\nu = \frac{1}{Z}e^{p_1\phi_1+\dots+p_k\phi_k}\,dx. \tag{22}$$

Then, one can check that for any $n \in \mathbb{N}$, $\delta > 0$, and Borel set A in $(\mathbb{R}^+)^n$,

$$\begin{aligned} e^{-2n(|p_1|+\dots+|p_k|)\delta}\frac{\mathrm{Leb}(A\cap C_n^\delta)}{\mathrm{Leb}(C_n^\delta)} &\le \nu^{\otimes n}\big(A|C_n^\delta\big) \\ &\le e^{2n(|p_1|+\dots+|p_k|)\delta}\frac{\mathrm{Leb}(A\cap C_n^\delta)}{\mathrm{Leb}(C_n^\delta)}. \end{aligned} \tag{23}$$

Thus, for any probability measure ν of the form (22),

$$\begin{aligned} &\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\nu^{\otimes\mathbb{N}}\big(A|C_n^\delta\big) \\ &= \limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log(\mathrm{Leb})^{\otimes\mathbb{N}}\big(A|C_n^\delta\big) \end{aligned} \tag{24}$$

and

$$\liminf_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log \nu^{\otimes\mathbb{N}}(A|C_n^{\delta}) = \liminf_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log(\mathrm{Leb})^{\otimes\mathbb{N}}(A|C_n^{\delta}).$$

This implies that Theorems 2.5 and 2.6 hold under general reference measures of type (22), particularly under the uniform distribution which is of our main interest. Also, the limiting law of the finite marginal distributions of (9) are identical under any reference measures of type (22) (see Remark 3.10).

On the other hand, the lower tail estimate of type (21) depends on the particular choice of the reference measure (22). This is because the scaling factor $g(n)$ in (21) grows slower than n at infinity. In fact, if we switch the reference measure from $(\nu_1)^{\otimes\mathbb{N}}$ to $(\nu_2)^{\otimes\mathbb{N}}$ for ν_1 and ν_2 of the form (22), then the cost arising from this change is $\mathcal{O}(e^{Cn\delta})$ in the sense that for some constant C,

$$e^{-nC\delta}\nu_2^{\otimes n}(A|C_n^{\delta}) \le \nu_1^{\otimes n}(A|C_n^{\delta}) \le e^{nC\delta}\nu_2^{\otimes n}(A|C_n^{\delta}).$$

Since the scaling factor $g(n)$ grows slowly than n at infinity, for any fixed $\delta > 0$,

$$\lim_{n\to\infty}\frac{1}{g(n)}\log(e^{Cn\delta}) = \infty.$$

This implies that the left-hand side of the lower tail estimate (21) is sensitive to the particular choice of the reference measure of the form (22).

Now, let us study the localization phenomenon by establishing the lower tail estimate (21). Since we have already proved in Theorem 2.6 that localization does not happen when $(a_1,\dots,a_{k-1}) \in \mathcal{S}_2$, we only consider the case $(a_1,\dots,a_{k-1}) \in \mathcal{S}_1$. Then, one can choose a (unique) probability measure ν on $(0,\infty)$ of the form

$$\nu = \frac{1}{Z}e^{p_1\phi_1+\cdots+p_{k-1}\phi_{k-1}}\,dx \tag{25}$$

(Z is a normalizing constant) satisfying

$$\int \phi_1\,d\nu = a_1,\dots,\int \phi_{k-1}\,d\nu = a_{k-1}$$

(see Lemma 3.12 for the explanations). Note that $\nu = \lambda^*$, which is the limiting distribution in Theorem 2.4 when $a_k \ge g_2(a_1,\dots,a_{k-1})$, satisfies this condition. Let us denote $1 \le m \le k-1$ by the largest index such that $p_m \ne 0$. As explained in Remark 2.7, the lower tail estimate (21) depends on the particular choice of the reference measure of form (22), and we will establish it under the reference measure $\mathbb{Q} := \nu^{\otimes\mathbb{N}}$.

The reason why we consider such reference measure to establish the lower tail estimate (21) is as follows. For the probability measure μ of the form (25), let us denote I^{μ} by the (weak) large deviation rate function for the sequence $(S_n^1,\dots,S_n^k)$ under $\mu^{\otimes\mathbb{N}}$. Then, when $(a_1,\dots,a_{k-1}) \in \mathcal{S}_1$ and $a_k > g_2(a_1,\dots,a_{k-1})$, due to the estimate (20) and Remark 2.7,

$$\mu^{\otimes\mathbb{N}}(C_n^{\delta}) = e^{-nI^{\mu}(a_1,\dots,a_k)+r_1(n,\delta)}$$

and

$$\mu^{\otimes\mathbb{N}}(\{M_n < a_k - g_2(a_1,\dots,a_{k-1}) - \epsilon\} \cap C_n^{\delta}) = e^{-nI^{\mu}(a_1,\dots,a_k)+r_2(n,\delta)}$$

for $r_1(n,\delta)$, $r_2(n,\delta)$ satisfying $\lim_{\delta\to 0}\lim_{n\to\infty}\frac{r_i(n,\delta)}{n} = 0$ for $i = 1,2$. In order to establish the lower tail estimate of type (21), we need to analyze the lower order terms $r_1(n,\delta)$, $r_2(n,\delta)$ since the scaling factor $g(n)$ grows slowly than n. Since the standard large deviation result does not reveal the finer behavior of $r_1(n,\delta)$ and $r_2(n,\delta)$, in order to capture this detailed structure we choose a probability measure μ such that $I^{\mu}(a_1,\dots,a_k)=0$. Since the probability measure ν chosen above satisfies $\int \phi_k\,d\nu = g_2(a_1,\dots,a_{k-1})$ (see Lemma 3.12), according to the law of large numbers and the estimate (98) in Lemma A.1, $I^{\nu}(a_1,\dots,a_k)=0$ whenever $a_k > g_2(a_1,\dots,a_{k-1})$.

As mentioned before, unlike the upper tail estimate (18) or the delocalization estimate (16), the scaling factor in the lower tail estimate (21) heavily relies on the structures of functions ϕ_i's in a complicated way. Roughly speaking, for a large class of functions ϕ_i's satisfying some technical conditions, the lower tail estimate (21) holds with the scaling factor $g(n) := (\phi_m \circ \phi_k^{-1})(n)$. We prove this in the particular case when $g(n)$ grows as n^{γ} $(0<\gamma<1)$ for the following two reasons: first of all, we try to keep arguments as simple as possible in order to separate the key ideas of the proof from technical details. Secondly, when ϕ_i's are polynomials, which is of our main interest due to its broad applications in geometry and PDE theory, $g(n)\approx n^{\gamma}$ for some $0<\gamma<1$. The following theorem provides the lower tail estimate for the maximum component, and describes the localization phenomenon:

THEOREM 2.8. *Suppose that* $(a_1,\dots,a_{k-1})\in \mathcal{S}_1$ *and* $a_k > g_2(a_1,\dots,a_{k-1})$. *Assume further that there exist* $0<\gamma_1,\dots,\gamma_{k-1}<1$ *such that for each* $1\le i\le k-1$,

$$\lim_{x\to\infty} \frac{(\phi_i\circ\phi_k^{-1})(x)}{x^{\gamma_i}} = 1. \tag{26}$$

If the reference measure $\mathbb{Q}$ *and the index* m *are chosen as above*, *then for any* $\epsilon>0$,

$$\limsup_{\delta\to 0}\limsup_{n\to\infty} \frac{1}{n^{\gamma_m}} \log \mathbb{Q}\big(M_n < a_k - g_2(a_1,\dots,a_{k-1}) - \epsilon \,|\, C_n^{\delta}\big) < 0. \tag{27}$$

In particular, localization happens in the sense that for any $\epsilon>0$,

$$\lim_{\delta\to 0}\lim_{n\to\infty} \mathbb{Q}\big(\big|M_n - \big(a_k - g_2(a_1,\dots,a_{k-1})\big)\big| < \epsilon \,|\, C_n^{\delta}\big) = 1. \tag{28}$$

Since the scaling factor in the lower tail estimate (27) grows slowly than n, we need a completely different approach from the standard large deviation theory to prove the estimate (27). In order to accomplish this, we partially adapt the method used in [2]. As mentioned before, following the proof of Theorem 2.8, one can check that for a large class of functions ϕ_i's satisfying some technical assumptions, the lower tail estimate (27) with the scaling factor $g(n) := (\phi_m\circ\phi_k^{-1})(n)$ holds as well.

It is important to note that Theorem 2.6 and Theorem 2.8 provide a complete picture of the localization and delocalization phenomena of the microcanonical ensembles with multiple constraints. In fact, let us assume that $(a_1,\dots,a_{k-1})\in\mathcal{S}_1$, and take the corresponding reference measure $\mathbb{Q}$ as in Theorem 2.8. If $a_k > g_2(a_1,\dots,a_{k-1})$, then the localization happens in the sense of (28), and the delocalization happens at all of the other sites (see Theorem 4.2 for details). On the other hand, if $a_k \le g_2(a_1,\dots,a_{k-1})$, then localization phenomenon does not happen according to Theorem 2.6 and Remark 2.7. Note that as mentioned before, when $(a_1,\dots,a_{k-1})\in\mathcal{S}_2$, whatever a_k is, localization phenomenon does not occur (see Theorem 2.6).

It is also crucial to note that when the localization happens, the maximum component M_n behaves differently in the upper tail and lower tail regime. In fact, the upper tail estimate is universal in the sense that the estimate (18) holds with the scaling factor n for any functions ϕ_i's satisfying Assumption 1. On the other hand, the lower tail estimate (27) is not universal in the sense that the scaling factor heavily relies on the structures of functions ϕ_i's. In the case when the localization does not happen (condition (i) or (ii) in Theorem 2.6), the delocalization estimate (16) is universal.

3. Large deviations and equivalence of ensembles results. In this section, we characterize the limit distribution to which the finite marginal distribution of (9) converges. As explained in Section 2, it exhibits a phase transition phenomenon. In Section 3.1, we briefly review the theory of large deviations and the classical equivalence of ensembles result. In Section 3.2 and 3.3, we prove Theorem 2.1 using a large deviation theory. In Section 3.4, we precisely characterize the limit distribution λ^* in Theorem 2.1 and conclude the proof of Theorem 2.4. Finally, in Section 3.5, we study a structure of the large deviation rate function for several empirical means.

3.1. *Preliminaries*: *Large deviation principle in statistical mechanics and the Gibbs conditioning principle*. The theory of large deviations has played an essential role in the equilibrium statistical mechanics. The sequence of probability distributions μ_n on the Polish space $\mathcal{S}$ are said to satisfy the *large deviation principle* (LDP) with the rate function I provided that for all Borel sets A,

$$
\begin{aligned}
-\inf_{x\in A^{\mathrm{o}}} I(x) &\le \liminf_{n\to\infty}\frac{1}{n}\log\mu_n(A^{\mathrm{o}}) \le \limsup_{n\to\infty}\frac{1}{n}\log\mu_n(\bar{A}) \\
&\le -\inf_{x\in\bar{A}} I(x).
\end{aligned}
$$

We say that *weak LDP* holds when the upper bound

$$
\limsup_{n\to\infty}\frac{1}{n}\log\mu_n(\bar{A}) \le -\inf_{x\in\bar{A}} I(x)
$$

holds only for compact sets $\bar{A}$. We require the rate function $I:\mathcal{S}\to[0,\infty]$ to be lower semicontinuous. I is said to be a *good* rate function if the set $\{x\in\mathcal{S}|I(x)\le c\}$ is compact for any $c\in\mathbb{R}$.

Once we have a large deviation principle for the sequence of probability distributions, we are able to study asymptotic behaviors of the conditional distributions. This can be rigorously stated as follows, which is called the *Gibbs conditioning principle* (see [12], Theorem 7.1).

THEOREM 3.1. *Let $\mathbb{P}_n$ be probability distributions on the Polish space $\mathcal{S}$ satisfying the large deviation principle with a good rate function I. Suppose that F and F_ϵ ($\epsilon>0$) are closed sets in $\mathcal{S}$ such that*:

(i) $I(F) := \inf_{x\in F} I(x) < \infty$.
(ii) $\mathbb{P}_n(F_\epsilon) > 0$ *for all n and $\epsilon > 0$.*
(iii) $F = \bigcap_{\epsilon>0} F_\epsilon$.
(iv) $F \subset (F_\epsilon)^{\mathrm{o}}$ *for all $\epsilon > 0$.*

Define M_F be a collection of $x\in F$ that minimize I over the set F. Then, for any open set G containing M_F,

$$
\limsup_{\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n(G^c|F_\epsilon) < 0. \tag{29}
$$

If in addition $M_F=\{x_0\}$ is a singleton, then

$$
\lim_{\epsilon\to 0}\lim_{n\to\infty}\mathbb{P}_n(\cdot|F_\epsilon) = \delta_{x_0}. \tag{30}
$$

Theorem 3.1 claims that as $n\to\infty$ followed by $\epsilon\to 0$, conditional distributions $\mathbb{P}_n(\cdot|F_\epsilon)$ get concentrated on the states at which the rate function is minimized over the constraint F.

Theorem 3.1 is a direct consequence of the definition of large deviation principle and the fact that

$$\inf_{x\in F} I(x) = \lim_{\epsilon\to 0}\inf_{x\in F_\epsilon} I(x), \tag{31}$$

which is true for a good rate function I (see [3], Lemma 4.1.6). However, in many cases, $\mathbb{P}_n$ only satisfy the weak LDP with a possibly nongood rate function. We now present a slightly generalized version of Theorem 3.1 which is useful in this paper.

THEOREM 3.2. *Let $\mathbb{P}_n$ be probability distributions on the Polish space $\mathcal{S}$ satisfying the weak large deviation principle with a rate function I. Assume that a set $\{I \le c\} \cap F_\epsilon$ is compact for any $c \in \mathbb{R}$ and $\epsilon > 0$. Then, the same conclusions* (29) *and* (30) *hold under the same assumptions* (i)–(iv) *in Theorem* 3.1.

PROOF. As in Theorem 3.1, it immediately follows from the definition of large deviation principle and the identity (31). We now show that (31) holds if $\{I \le c\} \cap F_\epsilon$ is compact for each $c \in \mathbb{R}$ and $\epsilon > 0$. Since $F \subset F_\epsilon$, it suffices to show that for any $\eta > 0$,

$$a := \lim_{\epsilon\to 0}\inf_{x\in F_\epsilon} I(x) \ge \inf_{x\in F} I(x) - \eta. \tag{32}$$

Since the set $\{I \le a+\eta\} \cap F_\epsilon$ is compact and closed,

$$\{I \le a+\eta\} \cap F = \bigcap_{\epsilon} \{\{I \le a+\eta\} \cap F_\epsilon\}$$

is also nonempty. This implies (32), which concludes the proof. □

As an application of the Gibbs conditioning principle, one can deduce the following classical result in the equilibrium statistical mechanics, which is called the principle of *equivalence of ensembles* (see [14], Chapter 5).

THEOREM 3.3. *Let $\lambda \in \mathcal{M}_1(\mathcal{S})$ and $\phi : \mathcal{S} \to \mathbb{R}$ be a bounded continuous function. Let us define $a := \lambda$-essinf ϕ and $b := \lambda$-esssup ϕ. For $\beta \in \mathbb{R}$, define $\mu_\beta \in \mathcal{M}_1(\mathcal{S})$ by $d\mu_\beta := \frac{1}{Z_\beta} e^{-\beta\phi}\, d\lambda$. Suppose that $\{X_k\}$'s are i.i.d. with individual distribution given by λ. Then, for $z \in (a,b)$, there exists a unique β such that*

$$\lim_{\delta\to 0^+}\lim_{n\to\infty} \mathbb{P}\left(X_1 \in \cdot \middle| \left|\frac{\phi(X_1)+\cdots+\phi(X_n)}{n} - z\right| \le \delta\right) = \mu_\beta.$$

Here, the inverse temperature β is chosen to satisfy $\int_{\mathcal{S}} \phi\, d\mu_\beta = z$.

It is not hard to check that similar result holds under the several constraints of type (6) with bounded and continuous observables ϕ_i's. We refer to [4] for the generalized version of Theorem 3.3, where the constraint is given by the bounded continuous interacting potentials. See also [7] for the case when the constraint is given by possibly unbounded interactions.

3.2. *Large deviations for the joint law of empirical distributions and several empirical means.* In this section, we obtain the large deviation results for the joint law of empirical distributions and several empirical means, which will play a crucial role in proving Theorem 2.1.

THEOREM 3.4. *Under the reference measure* $\mathbb{P}$, *the sequence* $(L_n, S_n^1, \dots, S_n^k)$ *in* $\mathcal{M}_1(\mathbb{R}^+) \times (\mathbb{R}^+)^k$ *satisfies the weak LDP with a rate function* J *given by*

$$J(\mu, v_1, \dots, v_k) = \begin{cases} H(\mu|\lambda) & \text{if } \int \phi_1 \, d\mu = v_1, \dots, \int \phi_{k-1} \, d\mu = v_{k-1}, \int \phi_k \, d\mu \le v_k, \\ \infty & \text{otherwise.} \end{cases}$$

This theorem will be crucially used to examine a thermodynamic behavior of the conditional distribution (9), with the aid of the Gibbs conditioning principle. In fact, once the weak LDP for the sequence $(L_n, S_n^1, \dots, S_n^k)$ is established, using Theorem 3.2, one can establish the asymptotic behavior of conditional distributions $\mathbb{P}(L_n \in \cdot | C_n^\delta)$ by solving a certain variational problem. This will be done in details in the rest of this section.

In view of Remark 2.7, one can also consider LDP under general reference measures $(\frac{1}{Z} e^{p_1\phi_1 + \dots + p_k\phi_k} \, dx)^{\otimes \mathbb{N}}$: a sequence $(L_n, S_n^1, \dots, S_n^k)$ enjoys the weak LDP with a rate function

$$J(\mu, v_1, \dots, v_k) = \begin{cases} -h(\mu) - p_1 v_1 - \dots - p_k v_k + C & \int \phi_i \, d\mu = v_i (1 \le i < k), \int \phi_k \, d\mu \le v_k, \\ \infty & \text{otherwise,} \end{cases}$$

for some constant C. Therefore, by the Gibbs conditioning principle, a particular choice of $p_1, \dots, p_k$ does not affect the large deviation asymptotic behavior of the conditional distributions $\mathbb{P}(L_n \in \cdot | C_n^\delta)$. This will be discussed in Remark 3.5. In Theorem 3.4, the reference measure $(\frac{1}{Z} e^{-\phi_1(x)} \, dx)^{\otimes \mathbb{N}}$ is considered.

PROOF. We follow the argument in [11]. We apply [3], Theorem 6.1.3, to obtain the weak LDP for the sequence $(L_n, S_n^1, \dots, S_n^k)$. This sequence is the empirical mean of the i.i.d. random variables $(\delta_{X_i}, \phi_1(X_i), \dots, \phi_k(X_i))$ taking values in $\mathcal{M}_1(\mathbb{R}^+) \times (\mathbb{R}^+)^k$. Let us denote $\mathcal{X} := \mathcal{M}(\mathbb{R}^+) \times \mathbb{R}^k$, which is equipped with the product topology of weak topology on the space of measures and the standard topology on $\mathbb{R}^k$, and similarly define $\mathcal{E} := \mathcal{M}_1(\mathbb{R}^+) \times (\mathbb{R}^+)^k$. It is not hard to check that Assumption 6.1.2 in [3] is satisfied in this setting (see [11], Lemma 3.2, for explanations in the case of $k = 1$). Thus, applying [3], Theorem 6.1.3, one can conclude that under the reference measure $\mathbb{P}$, the sequence $(L_n, S_n^1, \dots, S_n^k)$ satisfies the weak LDP with a rate function J given by

$$\begin{aligned} J(\mu, v_1, \dots, v_k) = \sup_{f \in C_b(\mathbb{R}^+), p_1, \dots, p_k \in \mathbb{R}} \bigg\{ & \int f \, d\mu + p_1 v_1 + \dots + p_k v_k \\ & - \log \int e^{f + p_1\phi_1 + \dots + p_k\phi_k} \, d\lambda \bigg\}. \end{aligned} \tag{33}$$

It is easy to check that for any function $f \in C_b(\mathbb{R}^+)$, $(p_1, \dots, p_k)$ satisfies

$$\int e^{f + p_1\phi_1 + \dots + p_k\phi_k} \, d\lambda < \infty$$

if and only if $(p_1, \dots, p_k)$ belongs to the set

$$\begin{aligned} D := & \{p_k < 0\} \cup \{p_k = 0, p_{k-1} < 0\} \cup \dots \cup \{p_k = \dots = p_3 = 0, p_2 < 0\} \\ & \cup \{p_k = \dots = p_3 = p_2 = 0, p_1 < 1\} \end{aligned} \tag{34}$$

thanks to the Assumption 1. Thus, it suffices to take the supremum over the set D in the expression (33). For each $(p_1, \dots, p_k) \in D$, let us define the auxiliary probability measure

$\nu^{(p_1,\dots,p_k)}$ on $(0,\infty)$ whose distribution is given by $\frac{1}{Z^{(p_1,\dots,p_k)}}e^{p_1\phi_1+\cdots+p_k\phi_k}\,d\lambda$ ($Z^{(p_1,\dots,p_k)}$ is a normalizing constant). Using the variation formula for the relative entropy,

$$H(\mu|\nu) = \sup_{f\in C_b}\left\{\int f\,d\mu - \log\int e^f\,d\nu\right\},$$

one can rewrite (33) as

$$\begin{aligned}
&J(\mu, v_1,\dots,v_k)\\
&\quad= \sup_{(p_1,\dots,p_k)\in D}\biggl\{p_1v_1+\cdots+p_kv_k - \log Z^{(p_1,\dots,p_k)}\\
&\qquad + \sup_{f\in C_b}\left(\int f\,d\mu - \log\int e^f\,d\nu^{(p_1,\dots,p_k)}\right)\biggr\}\\
&\quad= \sup_{(p_1,\dots,p_k)\in D}\{p_1v_1+\cdots+p_kv_k - \log Z^{(p_1,\dots,p_k)} + H(\mu|\nu^{(p_1,\dots,p_k)})\}\\
&\quad= \sup_{(p_1,\dots,p_k)\in D}\left\{p_1v_1+\cdots+p_kv_k + H(\mu|\lambda) - \int(p_1\phi_1+\cdots+p_k\phi_k)\,d\mu\right\}.
\end{aligned}$$

If we define the set $\mathcal{T}\subset\mathcal{M}_1(\mathbb{R}^+)$ by

$$\mathcal{T} := \left\{\mu\in\mathcal{M}_1(\mathbb{R}^+) : \int\phi_1\,d\mu = v_1,\dots,\int\phi_{k-1}\,d\mu = v_{k-1}, \int\phi_k\,d\mu\le v_k\right\},$$

then one can easily check that $J(\mu, v_1,\dots,v_k) = H(\mu|\lambda)$ when $\mu\in\mathcal{T}$ and ∞ otherwise. $\square$

REMARK 3.5. We present several remarks regarding Theorem 3.4.

1. If each function ϕ_i' is bounded and continuous, then it is obvious that the sequence $(L_n, S_n^1,\dots,S_n^k)$ satisfies the (full) LDP with a rate function J^{bounded} defined by

$$\begin{aligned}
&J^{\text{bounded}}(\mu, v_1,\dots,v_k)\\
&\quad=\begin{cases} H(\mu|\lambda) & \int\phi_1\,d\mu = v_1,\dots,\int\phi_k\,d\mu = v_k,\\ \infty & \text{otherwise.}\end{cases}
\end{aligned}$$

Theorem 3.4 implies that when ϕ_i's are unbounded functions satisfying Assumption 1, the rate function $J(\mu, v_1,\dots,v_k)$ may be finite even when $\int\phi_k\,d\mu\ne v_k$ since the weak topology induced on the space of probability measures is not strong enough to capture the behavior near the infinity. Note that $J(\mu, v_1,\dots,v_k)=\infty$ if $\int\phi_i\,d\mu\ne v_i$ for some $1\le i\le k-1$ since ϕ_k controls other functions $\phi_1,\dots,\phi_{k-1}$.

2. When we consider the pair of empirical distributions and a single empirical mean, the large deviation result Theorem 3.4 reads as follows (see [11], Lemma 3.3, in the case of $\phi(x)=x^p$ under the generalized Gaussian distribution): under the reference measure $(\frac{1}{Z}e^{-\phi}\,dx)^{\otimes\mathbb{N}}$, the sequence $(L_n, \frac{\phi(X_1)+\cdots+\phi(X_n)}{n})$ satisfies the (full) LDP with a good rate function

$$J(\mu, v) = \begin{cases} H(\mu|\lambda) + v - \int\phi\,d\mu & \text{if } \int\phi\,d\mu\le v,\\ \infty & \text{otherwise.}\end{cases}$$

3. For any fixed positive integer j, let us define

$$S_{n-j}^i := \frac{\phi_i(X_{j+1})+\cdots+\phi_i(X_n)}{n-j}$$

for each $1 \le i \le k$. Then, under the reference measure $\mathbb{P}$, the sequence $(L_n, S^1_{n-j}, \dots, S^k_{n-j})$ satisfies the weak LDP with the same rate function J defined in Theorem 3.4. Indeed, $(L_n, S^1_n, \dots, S^k_n)$ and $(L_n, S^1_{n-j}, \dots, S^k_{n-j})$ are exponentially equivalent sequences since for any realization,

$$\begin{aligned}\limsup_{n\to\infty} d\biggl(\frac{1}{n}(\delta_{X_1}+\cdots+\delta_{X_n}), \frac{1}{n-j}(\delta_{X_{j+1}}+\cdots+\delta_{X_n})\biggr)\\ \le \limsup_{n\to\infty}\frac{2j}{n}=0\end{aligned}$$

(d denotes the metric (7)).

4. For the probability measure $\mu \ll dx$ satisfying the condition

$$\int \phi_1\, d\mu = v_1, \dots, \int \phi_{k-1}\, d\mu = v_{k-1}, \int \phi_k\, d\mu \le v_k, \tag{35}$$

the rate function J in Theorem 3.4 can be written in terms of the differential entropy $h(\cdot)$: for some constant C,

$$\begin{aligned} H(\mu|\lambda) &= \int \log\biggl(\frac{d\mu}{d\lambda}\biggr)\, d\mu \\ &= \int \log\biggl(\frac{d\mu}{dx}\biggr)\, d\mu + \int \log\biggl(\frac{dx}{d\lambda}\biggr)\, d\mu \\ &= -h(\mu) + \int \phi_1\, d\mu + C = -h(\mu) + v_1 + C. \end{aligned} \tag{36}$$

In general, if the reference measure λ on $(0, \infty)$ is given by

$$\lambda = \frac{1}{Z} e^{p_1\phi_1+\cdots+p_k\phi_k}\, dx$$

for some $(p_1, \dots, p_k)$ for which the normalizing constant Z is finite, then by the same argument as in Theorem 3.4, the sequence $(L_n, S^1_n, \dots, S^k_n)$ under $\lambda^{\otimes\mathbb{N}}$ satisfies the weak LDP with the rate function J given by

$$\begin{aligned} &J(\mu, v_1, \dots, v_k) \\ &\quad = \begin{cases} H(\mu|\lambda) - p_k\Bigl(v_k - \int \phi_k\, d\mu\Bigr) & \int \phi_i\, d\mu = v_i\ (1 \le i \le k-1), \int \phi_k\, d\mu \le v_k, \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

For a probability measure $\mu \ll dx$ satisfying the condition (35), the rate function J can be written as

$$H(\mu|\lambda) - p_k\biggl(v_k - \int \phi_k\, d\mu\biggr) = -h(\mu) - p_1 v_1 - \cdots - p_k v_k + C.$$

Thus, in view of the Gibbs conditioning principle, conditional distributions $\mathbb{P}(L_n \in \cdot | C^\delta_n)$ weakly converge to the dirac mass at λ^* as $n \to \infty$ followed by $\delta \to 0$, where λ^* is a probability distribution maximizing the differential entropy $h(\mu)$ over the constraint (10) (see Section 3.3 for details). Therefore, the asymptotic behavior of $\mathbb{P}(L_n \in \cdot | C^\delta_n)$ does not depend on particular values of $p_1, \dots, p_k$.

We need the following simple lemma to ensure the existence and uniqueness of a minimizer of the relative entropy $H(\cdot|\lambda)$ over the set (10).

LEMMA 3.6. *The following set is closed, compact and convex*:

$$\mathcal{T} := \left\{ \mu \in \mathcal{M}_1(\mathbb{R}^+) : \int \phi_1 \, d\mu = v_1, \dots, \int \phi_{k-1} \, d\mu = v_{k-1}, \int \phi_k \, d\mu \le v_k \right\}.$$

PROOF. Suppose that $\mu_n \in \mathcal{T}$ and $\mu_n \to \mu$ as $n \to \infty$. We first show the closedness of $\mathcal{T}$ by proving that $\mu \in \mathcal{T}$. According to the Portmanteau theorem, $\int \phi_k \, d\mu \le v_k$ is obvious. Fix any $1 \le i \le k-1$ and let us show that $\int \phi_i \, d\mu = v_i$. Using Assumption 1 and the fact that $\int \phi_k \, d\mu_n \le v_k$, one can conclude that for any $\epsilon > 0$, there exists $M > 0$ such that for all n, $\int \phi_i \mathbb{1}_{[M,\infty)} \, d\mu_n < \epsilon$.

This implies that $\int \phi_i \mathbb{1}_{(0,M)} \, d\mu_n > v_i - \epsilon$. Since $\mu_n \to \mu$ and $\phi_i \mathbb{1}_{(0,M)} \in C_b(\mathbb{R}^+)$, we have $\lim_n \int \phi_i \mathbb{1}_{(0,M)} \, d\mu_n = \int \phi_i \mathbb{1}_{(0,M)} \, d\mu$. Therefore, we have $\int \phi_i \mathbb{1}_{(0,M)} \, d\mu \ge v_i - \epsilon$, and since ϵ is arbitrary, we obtain $\int \phi_i \, d\mu \ge v_i$. On the other hand, thanks to the Portmanteau theorem, $\int \phi_i \, d\mu \le v_i$. Thus, $\int \phi_i \, d\mu = v_i$, which concludes the closedness of $\mathcal{T}$.

Compactness of $\mathcal{T}$ immediately follows from the Prokhorov's theorem and Assumption 1. Convexity of $\mathcal{T}$ is also obvious. $\square$

According to the Sanov's theorem, the sequence of empirical distributions L_n satisfy the LDP with a rate function $H(\cdot|\lambda)$. Also, due to the generalized version of Cramér's theorem (see [3], Theorem 6.1.3), the sequence $(S_n^1, \dots, S_n^k)$ satisfies the weak LDP with a rate function $I(v_1, \dots, v_k)$ which is the Legendre transform of the logarithmic moment generating function:

$$H(p_1, \dots, p_k) = \log \int e^{p_1\phi_1 + \cdots + p_k\phi_k} \, d\lambda.$$

Since a map $\mu \to \int \phi_i \, d\mu$ may not be continuous, the rate function I cannot be directly obtained from the Sanov's theorem as a simple application of the standard contraction principle. However, applying Theorem 3.4, one can obtain the noncontinuous version of the contraction principle. It reveals the relation between two rate functions $H(\cdot|\lambda)$ and I.

PROPOSITION 3.7. *Under the reference measure $\mathbb{P}$, the sequence $(S_n^1, \dots, S_n^k)$ in $(\mathbb{R}^+)^k$ satisfies the weak LDP with a rate function $I(v_1, \dots, v_k)$ given by*

$$\begin{aligned} I(v_1, \dots, v_k) = \inf_{\mu \in \mathcal{M}_1(\mathbb{R}^+)} \Big\{ & H(\mu|\lambda) : \\ & \int \phi_1 \, d\mu = v_1, \dots, \int \phi_{k-1} \, d\mu = v_{k-1}, \int \phi_k \, d\mu \le v_k \Big\}. \end{aligned} \tag{37}$$

Also, $I(v_1, \dots, v_k)$ is the Legendre transform of $H(p_1, \dots, p_k)$ defined in (13).

PROOF. Let us apply the contraction principle to the projection $\pi : (L_n, S_n^1, \dots, S_n^k) \to (S_n^1, \dots, S_n^k)$. Since J is not necessarily a good rate function, in order that contraction principle works, we need to check that for $I(v_1, \dots, v_k)$ defined in (37),

$$\{(v_1, \dots, v_k) : I(v_1, \dots, v_k) \le c\} = \pi\big(\{(\mu, v_1, \dots, v_k) : J(\mu, v_1, \dots, v_k) \le c\}\big) \tag{38}$$

holds, and that this set is a closed set (see the proof of [3], Theorem 4.2.1). Note that since

$$\mathcal{T} := \left\{ \mu \in \mathcal{M}_1(\mathbb{R}^+) : \int \phi_1 \, d\mu = v_1, \dots, \int \phi_{k-1} \, d\mu = v_{k-1}, \int \phi_k \, d\mu \le v_k \right\}$$

is a closed set according to Lemma 3.6, the infimum of $H(\cdot|\lambda)$ is attained over $\mathcal{T}$ when $I(v_1, \dots, v_k) < \infty$. This implies that the equality in (38) holds. Also, since the sub-level set $\{H(\cdot|\lambda) \le c\}$ is compact with respect to the weak topology, under the projection π, the image of $\{(\mu, v_1, \dots, v_k) | J(\mu, v_1, \dots, v_k) \le c\}$ is closed. Therefore, the contraction principle is applicable, and the sequence $(S_n^1, \dots, S_n^k)$ satisfies the weak LDP with a rate function

$$I(v_1, \dots, v_k) := \inf_{\mu \in \mathcal{M}_1(\mathbb{R}^+)} J(\mu, v_1, \dots, v_k),$$

which immediately implies (37). Also, due to the uniqueness property of the rate function, the second part of proposition is obvious. □

REMARK 3.8. From the definition of the admissible set, for $(a_1, \dots, a_k) \in \mathcal{A}$,

$$\left\{\mu \in \mathcal{M}_1(\mathbb{R}^+) : h(\mu) \neq -\infty, \int \phi_i \, d\mu = a_i \ (1 \le i \le k-1), \int \phi_k \, d\mu \le a_k \right\}$$

is a nonempty set. Thus, according to Proposition 3.7, whenever $(a_1, \dots, a_k) \in \mathcal{A}$, $I(a_1, \dots, a_k) < \infty$ (see the identity (36)). This implies that for $(a_1, \dots, a_k) \in \mathcal{A}$, the microcanonical distribution $\mathbb{P}((X_1, \dots, X_n) \in \cdot | C_n^\delta)$ is well defined since for each $\delta > 0$, $\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(C_n^\delta) \ge -I(a_1, \dots, a_k) > -\infty$.

On the other hand, $I(a_1, \dots, a_k) = \infty$ when $a_k < g_1(a_1, \dots, a_{k-1})$.

Now, let us define $\lambda^* = \lambda^*(a_1, \dots, a_k)$ to be a unique minimizer of the relative entropy $H(\cdot|\lambda)$ over the set

$$\left\{\mu \in \mathcal{M}_1(\mathbb{R}^+) : \int \phi_1 \, d\mu = a_1, \dots, \int \phi_{k-1} \, d\mu = a_{k-1}, \int \phi_k \, d\mu \le a_k \right\}.$$

The existence and uniqueness of a minimizer follows from Lemma 3.6 and the lower semicontinuity, compact sublevel sets, strict convexity properties of the relative entropy functional $H(\cdot|\lambda)$. Note that λ^* is also a unique maximizer of the differential entropy $h(\cdot)$ due to the identity (36).

3.3. *Proof of Theorem* 2.1. In this section, we conclude the proof of Theorem 2.1. As an application of the Gibbs conditioning principle, combined with the large deviation result for the sequence $(L_n, S_n^1, \dots, S_n^k)$ obtained in Theorem 3.4, one can prove the following result.

LEMMA 3.9. *For any open set G containing λ^*,*

$$\limsup_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n \notin G | C_n^\delta) < 0.$$

PROOF. For each $0 < \delta < \min\{a_1, \dots, a_k\}$, define closed sets $F, F_\delta \subset \mathcal{M}_1(\mathbb{R}^+) \times (\mathbb{R}^+)^k$ by

$$F = \{(L_n, S_n^1, \dots, S_n^k) | S_n^1 = a_1, \dots, S_n^k = a_k\}$$

and

$$F_\delta = \{(L_n, S_n^1, \dots, S_n^k) | S_n^k \in [a_1 - \delta, a_1 + \delta], \dots, S_n^k \in [a_k - \delta, a_k + \delta]\}.$$

It is obvious that $F = \bigcap_{\delta>0} F_\delta$ and $F \subset (F_\delta)^{\mathrm{o}}$. Also, the infimum of $J(\mu, v_1, \dots, v_k)$ over the constraint $v_1 = a_1, \dots, v_k = a_k$ is attained at $(\mu, v_1, \dots, v_k) = (\lambda^*, a_1, \dots, a_k)$, and $G \times (\mathbb{R}^+)^k$ is an open neighborhood of $(\lambda^*, a_1, \dots, a_k)$. In addition, since sets $\{\mu : H(\mu|\lambda) \le c\}$ in $\mathcal{M}_1$ and $[a_1 - \delta, a_1 + \delta] \times \cdots \times [a_k - \delta, a_k + \delta]$ in $\mathbb{R}^k$ are compact, by the definition of a rate function J (see Theorem 3.4), $\{J \le c\} \cap F_\delta$ is compact in $\mathcal{M}_1 \times \mathbb{R}^k$. Therefore, according to the Gibbs conditioning principle Theorem 3.2,

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(L_n \notin G|C_n^\delta\big)$$
$$=\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big((L_n, S_n^1,\dots,S_n^k)\in G^c\times\mathbb{R}^k|C_n^\delta\big)<0. \qquad \square$$

As a corollary of the previous lemma, one can finish the proof of Theorem 2.1.

PROOF OF THEOREM 2.1 . Recall that a unique maximizer of the differential entropy $h(\cdot)$ over the set (10) coincides with a unique minimizer of the relative entropy $H(\cdot|\lambda)$ over the same set (10) (see the identity (36)). As a consequence of Lemma 3.9, we have

$$\lim_{\delta\to 0}\lim_{n\to\infty}\mathbb{P}\big(L_n\in\cdot|C_n^\delta\big)=\delta_{\lambda^*}.$$

According to [16], Proposition 2.2, this implies that for any fixed positive integer j,

$$\lim_{\delta\to 0}\lim_{n\to\infty}\mathbb{P}\big((X_1,\dots,X_j)\in\cdot|C_n^\delta\big)\to\big(\lambda^*\big)^{\otimes j}. \qquad \square$$

REMARK 3.10. Note that according to (24), Lemma 3.9 also holds under the uniform distribution on the constraint C_n^δ, which is of our main interest. Thus, the result in Theorem 2.1 holds under the uniform distribution as well.

3.4. *Characterization of the maximizer in Theorem 2.1*. In this section, we characterize the (unique) maximizer of the differential entropy $h(\cdot)$ over the set (10). Interestingly, it turns out that the maximizers have different forms in the case of $(a_1,\dots,a_{k-1})\in\mathcal{S}_1$ and $(a_1,\dots,a_{k-1})\in\mathcal{S}_2$. We first analyze the sets $\mathcal{S}_1$, $\mathcal{S}_2$ and the function g_2 defined in (15) in a more detailed way.

REMARK 3.11. For $(v_1,\dots,v_{k-1})\in\mathcal{S}_1$, there exist unique $p_1,\dots,p_{k-1}$ satisfying (14). This can be verified using the following facts:

- If $(v_1,\dots,v_{k-1},z)\in\partial H(p_1,\dots,p_{k-1},0)$ for some $p_1,\dots,p_{k-1}$, then for all $z<w$, $(v_1,\dots,v_{k-1},w)\in\partial H(p_1,\dots,p_{k-1},0)$.
- Rate function I is differentiable on $\mathcal{A}$.

Since $H(p_1,\dots,p_{k-1},p_k)=\infty$ for $p_k>0$, the first fact follows from the definition of the subdifferential of convex functions. The second fact immediately follows from the essentially strictly convexity of H and the fact that $\mathcal{A}\subset\mathrm{dom}(I)$. In fact, the essentially strictly convexity of H implies the essentially smoothness of I (see [15], Theorem 26.3). Since $\mathcal{A}\subset\mathrm{dom}(I)$ and $\mathcal{A}$ is open, the essentially smoothness of I implies that I is differentiable on $\mathcal{A}$.

Suppose that there exist $(p_1,\dots,p_{k-1})$ and $(p'_1,\dots,p'_{k-1})$ satisfying (14). Using the first fact above, there exists v_k such that $(v_1,\dots,v_k)\in\mathcal{A}$ and

$$(v_1,\dots,v_k)\in\partial H(p_1,\dots,p_{k-1},0),\ (v_1,\dots,v_k)\in\partial H\big(p'_1,\dots,p'_{k-1},0\big).$$

Since H is convex and lower semicontinuous, using the duality of H and I, we have $(p_1,\dots,p_{k-1},0),(p'_1,\dots,p'_{k-1},0)\in\partial I(v_1,\dots,v_{k-1},v_k)$. Since I is differentiable on $\mathcal{A}$, $(p_1,\dots,p_{k-1})$ satisfying (14) is unique.

The following lemma reveals useful properties of the sets $\mathcal{S}_1$, $\mathcal{S}_2$, and provides a formula for the function g_2.

LEMMA 3.12. *Suppose that* $(v_1,\dots,v_{k-1})\in\mathcal{S}_1$. *Then, for* $p_1,\dots,p_{k-1}$ *satisfying* (14),

$$v_i=\frac{1}{Z}\int\phi_i e^{p_1\phi_1+\cdots+p_{k-1}\phi_{k-1}}\,d\lambda \tag{39}$$

(*Z is a normalizing constant* $Z=\int e^{p_1\phi_1+\cdots+p_{k-1}\phi_{k-1}}\,d\lambda$) *for* $1\le i\le k-1$ *and*

$$g_2(v_1,\dots,v_{k-1})=\frac{1}{Z}\int\phi_k e^{p_1\phi_1+\cdots+p_{k-1}\phi_{k-1}}\,d\lambda. \tag{40}$$

Suppose that $(v_1,\dots,v_{k-1})\in\mathcal{S}_2$. *Then, for any* v_k *such that* $(v_1,\dots,v_k)\in\mathcal{A}$, *there exist* $p_1,\dots,p_k$ *such that* $p_k<0$ *and*

$$v_i=\frac{1}{Z}\int\phi_i e^{p_1\phi_1+\cdots+p_k\phi_k}\,d\lambda \tag{41}$$

(*Z is a normalizing constant* $Z=\int e^{p_1\phi_1+\cdots+p_k\phi_k}\,d\lambda$) *for* $1\le i\le k$.

PROOF. Let us consider the first case $(v_1,\dots,v_{k-1})\in\mathcal{S}_1$. By the definition of the set $\mathcal{S}_1$, there exist v_k and (unique) $p_1,\dots,p_{k-1}$ satisfying

$$(v_1,\dots,v_{k-1},v_k)\in\partial H(p_1,\dots,p_{k-1},0). \tag{42}$$

This implies that for any $\epsilon>0$, $H(p_1,\dots,p_{k-1},-\epsilon)-H(p_1,\dots,p_{k-1},0)\ge-\epsilon v_k$. Dividing this by $-\epsilon$ and then sending $\epsilon\to0^+$, using Fatou's lemma,

$$\frac{1}{Z}\int\phi_k e^{p_1\phi_1+\cdots+p_{k-1}\phi_{k-1}}\,d\lambda\le v_k. \tag{43}$$

Here, Z is a normalizing constant $Z=\int e^{p_1\phi_1+\cdots+p_{k-1}\phi_{k-1}}\,d\lambda$. This obviously implies that for any $1\le i\le k$,

$$w_i:=\frac{1}{Z}\int\phi_i e^{p_1\phi_1+\cdots+p_{k-1}\phi_{k-1}}\,d\lambda<\infty. \tag{44}$$

Using (42) again, for each $1\le i\le k-1$ and for any $\epsilon>0$, $c\in\mathbb{R}$, we have

$$\begin{aligned}&\lim_{\epsilon\to0^+}\frac{1}{\epsilon}\big[H(p_1,\dots,p_i+\epsilon c,\dots,p_{k-1},-\epsilon)-H(p_1,\dots,p_i,\dots,p_{k-1},0)\big]\\&\quad\ge cv_i-v_k.\end{aligned} \tag{45}$$

Using dominated convergence theorem, let us check that left-hand side of (45) is equal to cw_i-w_k. Indeed, if we denote A, A_ϵ ($\epsilon>0$) by

$$A:=e^{p_1\phi_1+\cdots+p_i\phi_i+\cdots+p_{k-1}\phi_{k-1}},\qquad A_\epsilon:=e^{p_1\phi_1+\cdots+(p_i+\epsilon c)\phi_i+\cdots+p_{k-1}\phi_{k-1}-\epsilon\phi_k},$$

then the left-hand side of (45) can be written as

$$\lim_{\epsilon\to0^+}\bigg[\frac{\log\int A_\epsilon\,d\lambda-\log\int A\,d\lambda}{\int A_\epsilon\,d\lambda-\int A\,d\lambda}\cdot\frac{\int A_\epsilon\,d\lambda-\int A\,d\lambda}{\epsilon}\bigg]. \tag{46}$$

Note that $\lim_{\epsilon\to0^+}\frac{A_\epsilon-A}{\epsilon}=\lim_{\epsilon\to0^+}A\cdot\frac{e^{\epsilon(c\phi_i-\phi_k)}-1}{\epsilon}=c\phi_iA-\phi_kA$. If we choose $M>0$ such that $x\ge M\Rightarrow c\phi_i(x)<\phi_k(x)$, then for $x\ge M$ and $\epsilon>0$, $|A\cdot\frac{e^{\epsilon(c\phi_i-\phi_k)}-1}{\epsilon}|\le A(\phi_k-c\phi_i)$. Also, if we denote $N:=\sup_{0<x\le M}|c\phi_i-\phi_k|<\infty$, then for $x\in(0,M)$ and $0<\epsilon<1$, $|A\cdot\frac{e^{\epsilon(c\phi_i-\phi_k)}-1}{\epsilon}|\le A(e^N-1)$. Note that $A(\phi_k-c\phi_i)\in L^1(d\lambda)$ due to (44), and $A\in L^1(d\lambda)$ since $(p_1,\dots,p_{k-1},0)\in\operatorname{dom}(H)$. Therefore, applying the dominated convergence theorem,

$$\lim_{\epsilon\to0^+}\int\frac{A_\epsilon-A}{\epsilon}\,d\lambda=\int(c\phi_i-\phi_k)e^{p_1\phi_1+\cdots+p_{k-1}\phi_{k-1}}\,d\lambda. \tag{47}$$

Also, since $\sup_{x\in(0,\infty)}(c\phi_i - \phi_k) < \infty$ and $A \in L^1(d\lambda)$, as an application of the dominated convergence theorem, one can deduce that $\lim_{\epsilon\to 0^+} \int A_\epsilon\, d\lambda = \int A\, d\lambda$. Thus,

$$\lim_{\epsilon\to 0^+} \frac{\log \int A_\epsilon\, d\lambda - \log \int A\, d\lambda}{\int A_\epsilon\, d\lambda - \int A\, d\lambda} = \left[\int e^{p_1\phi_1+\cdots+p_{k-1}\phi_{k-1}}\, d\lambda\right]^{-1}. \tag{48}$$

Using (46), (47) and (48), one can deduce that the left-hand side of (45) is equal to $cw_i - w_k$, and thus we have $cw_i - w_k \geq cv_i - v_k$. Since c is arbitrary, we obtain $w_i = v_i$, which implies (39). Also, using the convexity of H, it is easy to check that $(w_1, \dots, w_{k-1}, w_k) \in \partial H(p_1, \dots, p_{k-1}, 0)$. Since any v_k for which (42) holds satisfies (43), we obtain (40).

Finally, let us consider the case when $(v_1, \dots, v_{k-1}) \in \mathcal{S}_2$ and $(v_1, \dots, v_{k-1}, v_k) \in \mathcal{A}$. Since $(v_1, \dots, v_k) \in \mathcal{A} \subset \mathrm{int}(\mathrm{dom}(I))$ and I is essentially smooth, I is differentiable at $(v_1, \dots, v_k)$. If we choose $(p_1, \dots, p_k) \in \partial I(v_1, \dots, v_k)$, then by the Legendre duality, we have $(v_1, \dots, v_k) \in \partial H(p_1, \dots, p_k)$. Since $(v_1, \dots, v_{k-1}) \in \mathcal{S}_2$, $p_k \neq 0$. This in turn implies that $p_k < 0$ since $(p_1, \dots, p_k) \in \mathrm{dom}(\partial H) \subset \mathrm{dom}(H) = D$. Thus, H is differentiable at $(p_1, \dots, p_k)$, and we immediately obtain (41). □

Using Lemma 3.12, one can characterize a (unique) maximizer of the differential entropy $h(\cdot)$ over the set (10).

PROPOSITION 3.13. *Assume that λ^* is a unique maximizer of $h(\cdot)$ over the set (10). In the case of $(a_1, \dots, a_{k-1}) \in \mathcal{S}_1$ and $a_k \geq g_2(a_1, \dots, a_{k-1})$,*

$$\lambda^* = \frac{1}{Z} e^{p_1\phi_1+\cdots+p_{k-1}\phi_{k-1}}\, dx \tag{49}$$

for $p_1, \dots, p_{k-1}$ satisfying $\int \phi_i\, d\lambda^* = a_i$ for $1 \leq i \leq k-1$.

On the other hand, either in the case of

(i) $(a_1, \dots, a_{k-1}) \in \mathcal{S}_2$ *or*
(ii) $(a_1, \dots, a_{k-1}) \in \mathcal{S}_1$ *and* $a_k < g_2(a_1, \dots, a_{k-1})$,

$$\lambda^* = \frac{1}{Z} e^{p_1\phi_1+\cdots+p_k\phi_k}\, dx \tag{50}$$

for $p_1, \dots, p_k$ satisfying $p_k < 0$ and $\int \phi_i\, d\lambda^* = a_i$ for $1 \leq i \leq k$. In all cases, Z denotes the normalizing constant.

PROOF. Let us first consider the case $(a_1, \dots, a_{k-1}) \in \mathcal{S}_1$ and $a_k \geq g_2(a_1, \dots, a_{k-1})$. According to Lemma 3.12, there exists a probability measure ν of the form (49) satisfying $\int \phi_i\, d\nu = a_i$ for $1 \leq i \leq k-1$ and $\int \phi_k\, d\nu = g_2(a_1, \dots, a_{k-1}) \leq a_k$ (recall that $d\lambda$ is given by (8)). It is easy to check that ν is the maximizer of $h(\cdot)$ over the set (10). In fact, for any probability measure $\mu \ll dx$,

$$\begin{aligned} -h(\mu) &= H(\mu|\nu) + p_1 \int \phi_1\, d\mu + \cdots + p_{k-1} \int \phi_{k-1}\, d\mu + C \\ &\geq p_1 a_1 + \cdots + p_{k-1} a_{k-1} + C, \end{aligned}$$

and the equality is attained if and only if $\mu = \nu$.

Let us now consider the other cases, (i) and (ii). In each case, we first show the existence of a probability measure ν of the form (50) satisfying $\int \phi_i\, d\nu = a_i$ for $1 \leq i \leq k$. In the case of (i), it is already proved in Lemma 3.12, so we consider the case (ii). For $(p_1, \dots, p_k) \in \partial I(a_1, \dots, a_k)$, we have $(a_1, \dots, a_k) \in \partial H(p_1, \dots, p_k)$ by the Legendre duality. Since $a_k <$

$g_2(a_1,\dots,a_{k-1})$, we have $p_k \neq 0$, which in turn implies $p_k < 0$. This implies that H is differentiable at $(p_1,\dots,p_k)$, and for $1 \le i \le k$,

$$a_i = \frac{1}{Z}\int \phi_i e^{p_1\phi_1+\cdots+p_k\phi_k}\, d\lambda.$$

Now, as before, one can check that ν is the maximizer of $h(\cdot)$ over the set (10). In fact, since $p_k < 0$, for any probability measure $\mu \ll dx$,

$$\begin{aligned} -h(\mu) &= H(\mu|\nu) + p_1 \int \phi_1\, d\mu + \cdots + p_k \int \phi_k\, d\mu + C \\ &\ge p_1 a_1 + \cdots + p_k a_k + C, \end{aligned}$$

and the equality is attained if and only if $\mu = \nu$. $\square$

PROOF OF THEOREM 2.4. Theorem 2.1 and Proposition 3.13 immediately conclude the proof. $\square$

3.5. *Structure of the rate function I.* In this section, we establish useful properties of the rate function I. Recall that I is the weak LDP rate function for the sequence $(S_n^1,\dots,S_n^k)$ under the reference measure $\mathbb{P}$ (see Proposition 3.7). It turns out that $I(v_1,\dots,v_k)$ behaves differently when $(v_1,\dots,v_{k-1}) \in \mathcal{S}_1$ and $(v_1,\dots,v_{k-1}) \in \mathcal{S}_2$.

PROPOSITION 3.14. *For each $(v_1,\dots,v_{k-1}) \in (\mathbb{R}^+)^{k-1}$, the rate function $I(v_1,\dots,v_{k-1},\cdot)$ is nonincreasing. In the case of $(v_1,\dots,v_{k-1}) \in \mathcal{S}_1$,*

$$I(v_1,\dots,v_{k-1},z) > I(v_1,\dots,v_{k-1},w) \tag{51}$$

for all $z, w \in \mathbb{R}^+$ satisfying $z < w \le g_2(v_1,\dots,v_{k-1})$ and $(v_1,\dots,v_{k-1},w) \in \mathcal{A}$. Also, $I(v_1,\dots,v_{k-1},\cdot)$ is constant on the interval $[g_2(v_1,\dots,v_{k-1}),\infty)$.

On the other hand, in the case of $(v_1,\dots,v_{k-1}) \in \mathcal{S}_2$,

$$I(v_1,\dots,v_{k-1},z) > I(v_1,\dots,v_{k-1},w) \tag{52}$$

for all $z, w \in \mathbb{R}^+$ satisfying $z < w$ and $(v_1,\dots,v_{k-1},w) \in \mathcal{A}$.

PROOF. The proof consists of three steps.

Step 1. Nonincreasing property on $(0,\infty)$: recall the variational formula

$$I(v_1,\dots,v_k) = \sup_{(p_1,\dots,p_k)\in D} \big(p_1 v_1 + \cdots + p_k v_k - H(p_1,\dots,p_k)\big),$$

with the domain D defined in (34). For each $(p_1,\dots,p_k) \in D$, whenever $z < w$,

$$\begin{aligned} & p_1 v_1 + \cdots + p_{k-1} v_{k-1} + p_k z - H(p_1,\dots,p_k) \\ & \quad \ge p_1 v_1 + \cdots + p_{k-1} v_{k-1} + p_k w - H(p_1,\dots,p_k). \end{aligned}$$

Thus, $I(v_1,\dots,v_{k-1},z) \ge I(v_1,\dots,v_{k-1},w)$ when $z < w$.

Step 2. Case $(v_1,\dots,v_{k-1}) \in \mathcal{S}_1$: if $z < g_1(v_1,\dots,v_{k-1})$, then (51) is obvious since $I(v_1,\dots,v_{k-1},z) = \infty$ (see Remark 3.8). Now, assume that for some $g_1(v_1,\dots,v_{k-1}) \le z < w \le g_2(v_1,\dots,v_{k-1})$,

$$I(v_1,\dots,v_{k-1},z) = I(v_1,\dots,v_{k-1},w) < \infty.$$

Since $I(v_1,\dots,v_{k-1},\cdot)$ is nonincreasing, $I(v_1,\dots,v_{k-1},\cdot)$ is constant on the interval $[z,w]$. Thus, for any $y \in (z,w)$, the subgradient $(p_1,\dots,p_k)$ of I at $(v_1,\dots,v_{k-1},y)$ should satisfy $p_k = 0$. Since H and I are conjugate to each other, $(v_1,\dots,$

$v_{k-1}, y) \in \partial H(p_1, \dots, p_{k-1}, 0)$. This contradicts the definition of $g_2(v_1, \dots, v_{k-1})$ since $y < g_2(v_1, \dots, v_{k-1})$. Thus, (51) holds for z, w satisfying $z < w \le g_2(v_1, \dots, v_{k-1})$ and $(v_1, \dots, v_{k-1}, w) \in \mathcal{A}$.

Now, let us prove that $I(v_1, \dots, v_{k-1}, \cdot)$ is constant on the interval $[g_2(v_1, \dots, v_{k-1}), \infty)$. Due to the definition of g_2 and the fact (i) in Remark 3.11, for arbitrary $\epsilon > 0$, we have $(v_1, \dots, v_{k-1}, g_2(v_1, \dots, v_{k-1}) + \epsilon) \in \partial H(p_1, \dots, p_{k-1}, 0)$. This implies that

$$
\begin{aligned}
(53) \qquad & I\big(v_1, \dots, v_{k-1}, g_2(v_1, \dots, v_{k-1}) + \epsilon\big) + H(p_1, \dots, p_{k-1}, 0) \\
& \quad = \big(v_1, \dots, v_{k-1}, g_2(v_1, \dots, v_{k-1}) + \epsilon\big) \cdot (p_1, \dots, p_{k-1}, 0).
\end{aligned}
$$

Therefore, for any $x > 0$, using (53) and denoting $P := (p_1, \dots, p_{k-1}, 0)$,

$$
\begin{aligned}
& I\big(v_1, \dots, v_{k-1}, g_2(v_1, \dots, v_{k-1}) + x\big) \\
& \quad \ge \big(v_1, \dots, v_{k-1}, g_2(v_1, \dots, v_{k-1}) + x\big) \cdot P - H(P) \\
& \quad = \big(v_1, \dots, v_{k-1}, g_2(v_1, \dots, v_{k-1}) + \epsilon\big) \cdot P - H(P) \\
& \quad = I\big(v_1, \dots, v_{k-1}, g_2(v_1, \dots, v_{k-1}) + \epsilon\big).
\end{aligned}
$$

Since $x > 0$ is arbitrary and $I(v_1, \dots, v_{k-1}, \cdot)$ is nonincreasing, it follows from the above inequality that $I(v_1, \dots, v_{k-1}, \cdot)$ is constant on the interval $[g_2(v_1, \dots, v_{k-1}) + \epsilon, \infty)$. Since $\epsilon > 0$ is arbitrary and I is lower semicontinuous, $I(v_1, \dots, v_{k-1}, \cdot)$ is constant on the interval $[g_2(v_1, \dots, v_{k-1}), \infty)$.

Step 3. Case $(v_1, \dots, v_{k-1}) \in \mathcal{S}_2$: if $z < g_1(v_1, \dots, v_{k-1})$, then (52) is obvious since $I(v_1, \dots, v_{k-1}, z) = \infty$. Let us assume that for some $g_1(v_1, \dots, v_{k-1}) \le z < w$,

$$I(v_1, \dots, v_{k-1}, z) = I(v_1, \dots, v_{k-1}, w) < \infty.$$

Then, for any $y \in (z, w)$, the subgradient $(p_1, \dots, p_k)$ of I at $(v_1, \dots, v_{k-1}, y)$ should satisfy $p_k = 0$. Thus, by the Legendre duality, we have $(v_1, \dots, v_{k-1}, y) \in \partial H(p_1, \dots, p_{k-1}, 0)$, and this contradicts the definition of $\mathcal{S}_2$. Since we already proved the nonincreasing property of $I(v_1, \dots, v_{k-1}, \cdot)$, proof is concluded. $\square$

Proposition 3.14 will play a crucial role in analyzing the localization and delocalization phenomena of the microcanonical ensembles in Section 4.

4. Localization and delocalization of microcanonical ensembles. When the microcanonical ensemble is given by a single constraint, localization phenomenon does not happen in general (see Section 5.1 and Proposition 5.2 for details). However, when the microcanonical ensemble is given by multiple constraints, complicated localization behaviors can happen, as explained in Section 2. In this section, we systematically study the localization and delocalization phenomena of such ensembles using the theory of large deviations.

4.1. *Large deviations for the joint law of empirical distributions and the maximum component.* Let us define the maximum component M_n by

$$M_n := \frac{\max_{1 \le i \le n} \phi_k(X_i)}{n}.$$

The key ingredient that reveals the localization behavior is the large deviation result for the maximum component M_n. In order to capture the finer behavior of the microcanonical ensembles, we obtain a large deviation result for a sequence of joint law (L_n, M_n). As a consequence of the results developed in Section 3, we show that the joint law (L_n, M_n) enjoys a LDP with a rate function expressed in terms of J and I.

THEOREM 4.1. *For any Borel set A in $\mathcal{M}_1(\mathbb{R}^+) \times \mathbb{R}^+$,*

$$-\inf_{(\mu,z)\in A^{\mathrm{o}}} J_1^{\max}(\mu,z) \le \liminf_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big((L_n,M_n)\in A^{\mathrm{o}}|C_n^\delta\big)$$
$$\le \limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big((L_n,M_n)\in \bar{A}|C_n^\delta\big) \le -\inf_{(\mu,z)\in\bar{A}} J_1^{\max}(\mu,z)$$

with a rate function $J_1^{\max}$ given by

$$J_1^{\max}(\mu,z) := J(\mu,a_1,\dots,a_{k-1},a_k-z) - I(a_1,\dots,a_k). \tag{54}$$

Since we have formulas for the rate functions J and I,

$$J(\mu,v_1,\dots,v_k) = \begin{cases} H(\mu|\lambda) & \text{if } \int \phi_1\,d\mu = v_1,\dots,\int \phi_{k-1}\,d\mu = v_{k-1}, \int\phi_k\,d\mu \le v_k,\\ \infty & \text{otherwise.}\end{cases}$$

(see Theorem 3.4) and

$$I(v_1,\dots,v_k) = \inf_{\mu\in\mathcal{M}_1(\mathbb{R}^+)}\Big\{H(\mu|\lambda): \int\phi_1\,d\mu = v_1,\dots,\int\phi_{k-1}\,d\mu = v_{k-1},\int\phi_k\,d\mu\le v_k\Big\}$$

(see Proposition 3.7), one can express a rate function $J_1^{\max}$ in (54) in terms of the relative entropy. Recalling the fact that $H(\mu|\lambda) = -h(\mu) + \int\phi_1\,d\mu + C$ for some constant C, rate function $J_1^{\max}$ can be expressed in terms of the differential entropy and a function K

$$K(a_1,\dots,a_k) := \inf_{\mu\in\mathcal{M}_1(\mathbb{R}^+)}\Big\{-h(\mu):\int\phi_1\,d\mu = a_1,\dots,\int\phi_{k-1}\,d\mu = a_{k-1},\int\phi_k\,d\mu\le a_k\Big\},$$

which appear in Theorem 2.5. We first prove Theorem 4.1 using the results of Section 3, and then explain why $J_1^{\max}$ is equal to the rate function $J^{\max}$ in Theorem 2.5.

PROOF. Throughout this proof, we use the notations

$$S_{n-j}^i := \frac{\phi_i(X_{j+1})+\cdots+\phi_i(X_n)}{n-j},\ L_{n-j} := \frac{1}{n-j}(\delta_{X_{j+1}}+\cdots+\delta_{X_n})$$

for any fixed index j and $1\le i\le k$. Also, for $r>0$ and $\mu\in\mathcal{M}_1(\mathbb{R}^+)$, define $B(\mu,r)$, $\bar{B}(\mu,r)\subset\mathcal{M}_1(\mathbb{R}^+)$ by

$$B(\mu,r) := \{\nu: d(\nu,\mu)<r\},\qquad \bar{B}(\mu,r) := \{\nu: d(\nu,\mu)\le r\}.$$

Recall that d is a metric defined in (7) that induces the weak convergence of probability measures.

Step 1. Upper bound large deviations: it is obvious that

$$\mathbb{P}\big((L_n,M_n)\in\bar{A}|C_n^\delta\big) \le n\mathbb{P}\bigg(\Big(L_n,\frac{\phi_k(X_1)}{n}\Big)\in\bar{A}|C_n^\delta\bigg). \tag{55}$$

According to the LDP for the sequence $(S_n^1,\dots,S_n^k)$, for each $\delta>0$, we have

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}(C_n^\delta) \ge -\inf_{v_i\in(a_i-\delta,a_i+\delta)} I(v_1,\dots,v_k) \ge -I(a_1,\dots,a_k). \tag{56}$$

Let us define A^δ by a collection of $(\mu, y) \in \mathcal{M}_1(\mathbb{R}^+) \times \mathbb{R}^+$ for which there exists $x \in \mathbb{R}^+$ satisfying $(\mu, x) \in \bar{A}$ and $|y - (a_k - x)| < \delta$. Then, using the condition (iii) in Assumption 1, for sufficiently large n,

$$\left\{\left(L_n, \frac{\phi_k(X_1)}{n}\right) \in \bar{A}\right\} \cap C_n^\delta \quad \Rightarrow \quad B_n^\delta := \bigcap_{i=1}^{k-1} \{|S_{n-1}^i - a_i| < 2\delta\} \cap \{(L_n, S_{n-1}^k) \in A^\delta\}.$$

According to the LDP result Theorem 3.4 and Remark 3.5,

$$\tag{57} \begin{aligned} &\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big((L_n, S_{n-1}^1 \cdots, S_{n-1}^k) \in B_n^\delta\big) \\ &\quad \le - \inf_{(\mu, v_k)\in A^\delta, v_1 \in [a_1 - 2\delta, a_1 + 2\delta], \dots, v_{k-1} \in [a_{k-1} - 2\delta, a_{k-1} + 2\delta]} J(\mu, v_1, \dots, v_k) \end{aligned}$$

Note that since the sequence $\{L_n\}$ under $\mathbb{P}$ is exponentially tight and $\prod_{i=1}^{k-1}[a_i - 2\delta, a_i + 2\delta] \times [0, a_k + \delta]$ is compact, the weak LDP result Theorem 3.4 is applicable. Sending $\delta \to 0$, using [3], Lemma 4.1.6,

$$\tag{58} \begin{aligned} &\lim_{\delta\to 0} \inf_{(\mu, v_k)\in A^\delta, v_1 \in [a_1 - 2\delta, a_1 + 2\delta], \dots, v_{k-1} \in [a_{k-1} - 2\delta, a_{k-1} + 2\delta]} J(\mu, v_1, \dots, v_k) \\ &\quad = \inf_{(\mu, v_k)\in \bar{A}} J(\mu, a_1, \dots, a_{k-1}, a_k - z). \end{aligned}$$

Note that although J is not necessarily a good rate function, [3], Lemma 4.1.6, is applicable since intervals $[a_i - 2\delta, a_i + 2\delta]$ and $[0, a_k + \delta]$ are compact and the relative entropy has compact sub-level sets. Therefore, using (56), (57), and (58),

$$\begin{aligned} &\limsup_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left(L_n, \frac{\phi_k(X_1)}{n}\right) \in \bar{A} | C_n^\delta\right) \\ &\quad \le \limsup_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left\{\left(L_n, \frac{\phi_k(X_1)}{n}\right) \in \bar{A}\right\} \cap C_n^\delta\right) - \liminf_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(C_n^\delta) \\ &\quad \le - \inf_{(\mu, z)\in \bar{A}} J_1^{\max}(\mu, z). \end{aligned}$$

This and (55) conclude the proof of upper bound large deviation.

Step 2. Lower bound large deviations: it suffices to show that for any $z, \epsilon > 0$ and open set U containing arbitrary $\mu \in \mathcal{M}_1(\mathbb{R}^+)$,

$$\tag{59} -J_1^{\max}(\mu, z) \le \liminf_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big((L_n, M_n) \in U \times (z - \epsilon, z + \epsilon) | C_n^\delta\big).$$

If $\int \phi_k \, d\mu > a_k - z$ or $\int \phi_i \, d\mu \neq a_i$ for some $1 \le i \le k - 1$, then (59) is obvious since $J_1^{\max}(\mu, z) = \infty$. Thus, throughout the proof we assume that $\int \phi_k \, d\mu \le a_k - z$ and $\int \phi_i \, d\mu = a_i$ for $1 \le i \le k - 1$. Since ϕ_k is bounded from below and continuous, according to the Portmanteau theorem, there exists $r_0 > 0$ such that

$$\tag{60} d(\nu, \mu) < r_0 \quad \Rightarrow \quad \int \phi_k \, d\nu > \int \phi_k \, d\mu - z.$$

Take a positive integer $j \ge 2$ such that $a_k - jz < \int \phi_k \, d\mu \le a_k - (j - 1)z$, and denote $0 \le w := a_k - (j - 1)z - \int \phi_k \, d\mu < z$. Also, define two events $E_{n,\delta}^1$ and E_n^2 by

$$\begin{aligned} E_{n,\delta}^1 &:= \bigcap_{i=1}^{j-1} \left\{\left|\frac{\phi_k(X_i)}{n} - z\right| \le \frac{\delta}{4(j-1)}\right\} \cap \left\{\left|\frac{\phi_k(X_j)}{n} - w\right| \le \frac{\delta}{4}\right\}, \\ E_n^2 &:= \bigcap_{i=j+1}^{n} \left\{\frac{\phi_k(X_i)}{n} < z\right\}. \end{aligned}$$

It is obvious that for sufficiently small $\delta > 0$,

$$E^1_{n,\delta} \cap E^2_n \quad \Rightarrow \quad M_n \in (z-\epsilon, z+\epsilon).$$

Therefore, for the open set $U = B(\mu, r)$ with $r < \frac{r_0}{2}$, for sufficiently small $\delta > 0$,

$$\begin{aligned} &\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big((L_n, M_n) \in U \times (z-\epsilon, z+\epsilon) | C^\delta_n\big) \\ &\quad \geq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(\{L_n \in U\} \cap E^1_{n,\delta} \cap E^2_n | C^\delta_n\big) \\ &\quad \geq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(\{L_n \in U\} \cap E^1_{n,\delta} \cap E^2_n \cap C^\delta_n\big) - \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(C^\delta_n\big). \end{aligned} \tag{61}$$

According to [3], Lemma 4.1.6,

$$\begin{aligned} &\limsup_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(C^\delta_n\big) \\ &\quad \leq \limsup_{\delta\to 0} \Big[- \inf_{v_1 \in [a_1-\delta, a_1+\delta], \ldots, v_k \in [a_k-\delta, a_k+\delta]} I(v_1, \ldots, v_k) \Big] \\ &\quad = -I(a_1, \ldots, a_k). \end{aligned} \tag{62}$$

Also, it is obvious that

$$\begin{aligned} &\mathbb{P}\big(\{L_n \in U\} \cap E^1_{n,\delta} \cap E^2_n \cap C^\delta_n\big) \\ &\quad = \mathbb{P}\big(\{L_n \in U\} \cap E^1_{n,\delta} \cap C^\delta_n\big) - \mathbb{P}\big(\{L_n \in U\} \cap E^1_{n,\delta} \cap \big(E^2_n\big)^c \cap C^\delta_n\big) \\ &\quad \geq \mathbb{P}\big(\{L_n \in U\} \cap E^1_{n,\delta} \cap C^\delta_n\big) \\ &\qquad - (n-j)\mathbb{P}\bigg(\{L_n \in U\} \cap E^1_{n,\delta} \cap \bigg\{\frac{\phi_k(X_{j+1})}{n} \geq z\bigg\} \cap C^\delta_n\bigg). \end{aligned} \tag{63}$$

Let us first estimate the following quantity:

$$\liminf_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(\big\{L_n \in B(\mu, r)\big\} \cap E^1_{n,\delta} \cap C^\delta_n\big).$$

For sufficiently small $\delta > 0$, one can take open sets D^δ_n in $(\mathbb{R}^+)^k$ such that for sufficiently large n,

$$\prod_{i=1}^{k-1} \left(a_i - \frac{\delta}{2}, a_i + \frac{\delta}{2}\right) \times \left(\int \phi_k \, d\mu - \frac{\delta}{2}, \int \phi_k \, d\mu + \frac{\delta}{2}\right) \subset D^\delta_n$$

and

$$E^1_{n,\delta} \cap \big\{\big(S^1_{n-j}, \ldots, S^{k-1}_{n-j}, S^k_{n-j}\big) \in D^\delta_n\big\} \quad \Rightarrow \quad E^1_{n,\delta} \cap C^\delta_n$$

thanks to the condition (iii) in Assumption 1. Therefore, according to the LDP result Theorem 3.4, for sufficiently small $\delta > 0$,

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\{L_n \in B(\mu, r)\} \cap E^1_{n,\delta} \cap C^\delta_n) \\
&\quad \geq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left\{L_{n-j} \in B\left(\mu, \frac{r}{2}\right)\right\} \cap E^1_{n,\delta} \cap \{(S^1_{n-j}, \dots, S^k_{n-j}) \in D^\delta_n\}\right) \\
&\quad \geq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(E^1_{n,\delta}) \\
&\qquad + \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left((L_{n-j}, S^1_{n-j}, \dots, S^k_{n-j}) \in B\left(\mu, \frac{r}{2}\right) \times D^\delta_n\right) \qquad (64)\\
&\quad \geq - \inf_{\nu \in B(\mu, \frac{r}{2}), (v_1, \dots, v_k) \in D^\delta_n} J(\nu, v_1, \dots, v_k) \\
&\quad \geq -J\left(\mu, a_1, \dots, a_{k-1}, \int \phi_k \, d\mu\right) \\
&\quad = -H(\mu|\lambda) = -J(\mu, a_1, \dots, a_{k-1}, a_k - z).
\end{aligned}$$

Note that in the fourth line, we used (98) in Lemma A.1.

Now, let us show that

$$\limsup_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\{L_n \in B(\mu, r)\} \cap E^1_{n,\delta} \cap \left\{\frac{\phi_k(X_{j+1})}{n} \geq z\right\} \cap C^\delta_n\right) = -\infty. \qquad (65)$$

Note that under $E^1_{n,\delta} \cap \{\frac{\phi_k(X_{j+1})}{n} \geq z\} \cap C^\delta_n$,

$$a_k - \int \phi_k \, d\mu + z - \frac{\delta}{2} < \frac{\phi_k(X_1) + \cdots + \phi_k(X_{j+1})}{n} \leq a_k + \delta,$$

which implies that $\int \phi_k \, d\mu > z - \frac{3\delta}{2}$. Thus, if $\int \phi_k \, d\mu < z$, then $E^1_{n,\delta} \cap \{\frac{\phi_k(X_{j+1})}{n} \geq z\} \cap C^\delta_n$ is an empty set for sufficiently small $\delta > 0$, which implies (65). Thus, from now on we assume that $\int \phi_k \, d\mu \geq z$. One can take closed sets F^δ_n in $(\mathbb{R}^+)^k$ such that for sufficiently large n,

$$F^\delta_n \subset \prod_{i=1}^{k-1} [a_i - 2\delta, a_i + 2\delta] \times \left[0, \int \phi_k \, d\mu - z + 2\delta\right]$$

and

$$E^1_{n,\delta} \cap \left\{\frac{\phi_k(X_{j+1})}{n} \geq z\right\} \cap C^\delta_n \quad \Rightarrow \quad (S^1_{n-j-1}, \dots, S^k_{n-j-1}) \in F^\delta_n.$$

Applying the LDP result Theorem 3.4 and Remark 3.5, we have

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\{L_n \in B(\mu, r)\} \cap E^1_{n,\delta} \cap \left\{\frac{\phi_k(X_{j+1})}{n} \geq z\right\} \cap C^\delta_n\right) \\
&\quad \leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}((L_n, S^1_{n-j-1}, \dots, S^k_{n-j-1}) \in \bar{B}(\mu, r) \times F^\delta_n) \qquad (66)\\
&\quad \leq - \inf_{\nu \in \bar{B}(\mu, r), (v_1, \dots, v_k) \in F^\delta_n} J(\nu, v_1, \dots, v_k).
\end{aligned}$$

Taking a limit $\delta \to 0$, using [3], Lemma 4.1.6,

$$\lim_{\delta\to 0} \inf_{\nu \in \bar{B}(\mu, r), (v_1, \dots, v_k) \in F^\delta_n} J(\nu, v_1, \dots, v_k) = \inf_{\nu \in \bar{B}(\mu, r), v_k \in [0, \int \phi_k \, d\mu - z]} J(\nu, a_1, \dots, a_{k-1}, v_k).$$

Since we chose r_0 satisfying (60) and $r < \frac{r_0}{2}$,

$$\inf_{\nu \in \bar{B}(\mu,r), v_k \in [0, \int \phi_k \, d\mu - z]} J(\nu, a_1, \dots, a_{k-1}, v_k) = -\infty.$$

Therefore, sending $\delta \to 0$ in (66), one can deduce (65). Applying (65) and (64) to (63), we obtain

$$\begin{aligned} &\liminf_{\delta \to 0} \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}\big(\{L_n \in B(\mu, r)\} \cap E^1_{n,\delta} \cap E^2_{n,\delta} \cap C^\delta_n\big) \\ &\quad \geq -J(\mu, a_1, \dots, a_k - z). \end{aligned} \tag{67}$$

Thus, using (61), (62) and (67), we finally obtain (59) since

$$\begin{aligned} &\liminf_{\delta \to 0} \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}\big((L_n, M_n) \in B(\mu, r) \times (z - \epsilon, z + \epsilon) | C^\delta_n\big) \\ &\geq -J(\mu, a_1, \dots, a_{k-1}, a_k - z) + I(a_1, \dots, a_k) \\ &= -J^{\max}_1(\mu, z). \end{aligned}$$

□

PROOF OF THEOREM 2.5. Recall that when $\mu \ll dx$, $H(\mu|\lambda) = -h(\mu) + \int \phi_1 \, d\mu + C$ for some constant C. Thus, using Proposition 3.7 and the rate function formula for J in Theorem 3.4, one can conclude that

$$\begin{aligned} &J^{\max}_1(\mu, z) \\ &\quad = \begin{cases} -h(\mu) - K(a_1, \dots, a_k) & \int \phi_i \, d\mu = a_i \ (1 \leq i \leq k-1), \int \phi_k \, d\mu \leq a_k - z, \\ \infty & \text{otherwise,} \end{cases} \end{aligned}$$

for

$$\begin{aligned} &K(a_1, \dots, a_k) \\ &\quad = \inf_{\mu \in \mathcal{M}_1(\mathbb{R}^+)} \left\{ -h(\mu) : \int \phi_1 \, d\mu = a_1, \dots, \int \phi_{k-1} \, d\mu = a_{k-1}, \int \phi_k \, d\mu \leq a_k \right\}. \end{aligned}$$

This concludes the proof of Theorem 2.5. □

4.2. *Localization and delocalization.* In this section, we study the localization and delocalization phenomena using the large deviation result Theorem 2.5. First, we prove Theorem 2.6, which is about the delocalization result.

PROOF OF THEOREM 2.6. First, let us consider the case when $(a_1, \dots, a_{k-1}) \in \mathcal{S}_1$ and $a_k > g_2(a_1, \dots, a_{k-1})$. Applying the LDP result Theorem 4.1 and Proposition 3.7,

$$\begin{aligned} &\limsup_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\big(M_n \in [a_k - g_2(a_1, \dots, a_{k-1}) + \epsilon, a_k] | C^\delta_n\big) \\ &\quad \leq -\inf_{z \in [a_k - g_2(a_1, \dots, a_{k-1}) + \epsilon, a_k]} I(a_1, \dots, a_{k-1}, a_k - z) + I(a_1, \dots, a_k) \\ &\quad = -\inf_{w \in [0, g_2(a_1, \dots, a_{k-1}) - \epsilon]} I(a_1, \dots, a_{k-1}, w) + I(a_1, \dots, a_k) < 0. \end{aligned}$$

The last inequality follows from Proposition 3.14.

Now, suppose that (i) or (ii) holds. Applying the LDP result Theorem 4.1 and Proposition 3.7 again, we have

$$\begin{aligned}
&\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(M_n\in[\epsilon,a_k]|C_n^\delta\big)\\
&\quad\le-\inf_{z\in[\epsilon,a_k]} I(a_1,\dots,a_{k-1},a_k-z)+I(a_1,\dots,a_k)\\
&\quad=-\inf_{w\in[0,a_k-\epsilon]} I(a_1,\dots,a_{k-1},w)+I(a_1,\dots,a_k)<0.
\end{aligned}$$

The last inequality follows from Proposition 3.14. □

We have shown that when $(a_1,\dots,a_k)$ satisfies (i) or (ii) in Theorem 2.6, localization does not happen. We now consider the case when $(a_1,\dots,a_{k-1})\in\mathcal{S}_1$ and $a_k>g_2(a_1,\dots,a_{k-1})$. As explained in Section 2, unlike the upper tail estimate (18) for the maximum component M_n, the lower tail estimate

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(M_n\le a_k-g_2(a_1,\dots,a_{k-1})-\epsilon|C_n^\delta\big)<0 \tag{68}$$

does not hold. In fact, according to the large deviation result Theorem 4.1 and Proposition 3.14, we have

$$\begin{aligned}
&\liminf_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(M_n<a_k-g_2(a_1,\dots,a_{k-1})-\epsilon|C_n^\delta\big)\\
&\quad\ge-\inf_{z\in[0,a_k-g_2(a_1,\dots,a_{k-1})-\epsilon)} I(a_1,\dots,a_{k-1},a_k-z)+I(a_1,\dots,a_{k-1},a_k)=0.
\end{aligned}$$

As mentioned in Section 2, unlike the upper tail estimate (18), the correct scaling factor in the lower tail estimate of type (68) highly depends on the structures of functions ϕ_i's. We now prove Theorem 2.8, which is about the lower tail estimate and the localization result. Since the correct scaling factor grows slowly than n, the proof is completely different from the standard large deviation arguments we have used so far, and we partially adapt the idea in [2].

PROOF OF THEOREM 2.8. We partially follow the argument in [2]. Recall that $1\le m\le k-1$ is the largest index such that $p_m\neq 0$, and it is obvious that $p_m<0$. Throughout the proof, we define $s:=a_k-g_2(a_1,\dots,a_{k-1})$ and choose a sufficiently small $\theta>0$ such that $p_m+3\theta<0$. In order to alleviate the notation, we define $\gamma:=\gamma_m$. Choose two numbers $0<\alpha,\beta<1$ satisfying

$$\frac{1}{2}(1+\gamma+2\alpha)<\beta<1. \tag{69}$$

We first compute the lower bound of $\liminf_{n\to\infty}\frac{1}{n^\gamma}\log\mathbb{Q}(C_n^\delta)$. It is obvious that

$$\begin{aligned}
&\bigcap_{i=1}^{k-1}\Big\{|S_{n-1}^i-a_i|<\frac{\delta}{2}\Big\}\cap\Big\{|S_{n-1}^k-g_2(a_1,\dots,a_{k-1})|<\frac{\delta}{2}\Big\}\cap\Big\{\Big|\frac{\phi_k(X_1)}{n}-s\Big|<\frac{\delta}{2}\Big\}\\
&\quad\Rightarrow\quad C_n^\delta.
\end{aligned}$$

Since $\int\phi_i\,d\nu=a_i$ for $1\le i\le k-1$ and $\int\phi_k\,d\nu=g_2(a_1,\dots,a_{k-1})$ (see Lemma 3.12), according to the law of large numbers,

$$\lim_{n\to\infty}\mathbb{Q}\bigg(\bigcap_{i=1}^{k-1}\Big\{|S_{n-1}^i-a_i|<\frac{\delta}{2}\Big\}\cap\Big\{|S_{n-1}^k-g_2(a_1,\dots,a_{k-1})|<\frac{\delta}{2}\Big\}\bigg)=1. \tag{70}$$

Thus, combining (100) in Lemma A.1 with (70), we obtain

$$\liminf_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n^\gamma}\log\mathbb{Q}(C_n^\delta)\geq p_m s^\gamma. \tag{71}$$

Now, let us compute the upper bound of

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n^\gamma}\log\mathbb{Q}(\{M_n<s-\epsilon\}\cap C_n^\delta).$$

For $n\in\mathbb{N}$, let us choose $f(n)$ satisfying $\phi_k(f(n))=n^\alpha$, and define $u_n:=\mathbb{E}^\nu[\phi_k(X_i)\times\mathbb{1}_{X_i\leq f(n)}]$. Note that f is increasing and $\lim_{n\to\infty}f(n)=\infty$. Using Assumption 1, the condition (26), and the change of variables, for sufficiently large n,

$$\begin{aligned}g_2(a_1,\dots,a_{k-1})-u_n&=\int_{f(n)}^\infty\phi_k\,d\nu\\&\leq\int_{f(n)}^\infty\phi_k e^{(p_m+\theta)\phi_m}\,dx\\&\leq\int_{n^\alpha}^\infty ye^{(p_m+2\theta)y^\gamma}y^M\,dy\leq C\exp((p_m+3\theta)n^{\alpha\gamma}).\end{aligned}\tag{72}$$

Define the event E_n^1 by

$$E_n^1:=\left\{\left|\sum_{i=1}^n(\phi_k(X_i)\mathbb{1}_{X_i\leq f(n)}-u_n)\right|>n^\beta\right\}.$$

Since $0\leq\phi_k(X_i)\mathbb{1}_{X_i\leq f(n)}\leq n^\alpha$, according to the Hoeffding's inequality [10],

$$\mathbb{Q}(E_n^1)\leq 2\exp(-2n^{2\beta-1-2\alpha}).\tag{73}$$

Now, define E_n^2 to be the event for which there exists the set of indices I satisfying $|I|=h(n):=[n^{\gamma-\frac{\alpha\gamma}{2}}]$ such that $X_i>f(n)$ for all $i\in I$. Then, using Assumption 1 and the change of variables, for sufficiently large n,

$$\begin{aligned}\mathbb{Q}(E_n^2)&<\binom{n}{h(n)}\left[\int_{f(n)}^\infty e^{(p_m+\theta)\phi_m}\,dx\right]^{h(n)}\\&<Cn^{h(n)}\left[\int_{n^\alpha}^\infty e^{(p_m+2\theta)y^\gamma}y^M\,dy\right]^{h(n)}<C\exp[C(p_m+3\theta)n^{\gamma+\frac{\alpha\gamma}{2}}].\end{aligned}\tag{74}$$

Finally, let us fix a constant $\eta>0$ satisfying

$$s-\epsilon<\left(\frac{s}{(s+\eta)^\gamma}\right)^{\frac{1}{1-\gamma}},\tag{75}$$

and then define E_n^3 to be the event for which $\sum_{i\in I}\phi_m(X_i)>(s+\eta)^\gamma n^\gamma$ for some I satisfying $|I|<h(n)$. Using the result (99) in Lemma A.1, for sufficiently large n,

$$\mathbb{Q}(E_n^3)<C\exp[(p_m+2\theta)((s+\eta)^\gamma n^\gamma-Ch(n))].\tag{76}$$

Now, let us check that

$$(E_n^1)^c\cap(E_n^2)^c\cap(E_n^3)^c\cap C_n^\delta\quad\Rightarrow\quad\left\{M_n>\left(\frac{s-2\delta}{(s+\eta)^\gamma}\right)^{\frac{1}{1-\gamma}}\right\}\cap C_n^\delta.\tag{77}$$

If we define $I:=\{1\leq i\leq n|X_i>f(n)\}$, then $(E_n^2)^c\cap(E_n^3)^c$ imply $|I|<h(n)$ and

$$\sum_{i\in I}\phi_m(X_i)\leq(s+\eta)^\gamma n^\gamma.\tag{78}$$

Under the event $(E_n^1)^c \cap C_n^\delta$, $|\sum_{i\in I}\phi_k(X_i) - (a_k - u_n)n| < \delta n + n^\beta$. Combining this with (72), we obtain

$$
\begin{aligned}
\Big|\sum_{i\in I}\phi_k(X_i) - sn\Big| &= \Big|\sum_{i\in I}\phi_k(X_i) - \big(a_k - g_2(a_1,\dots,a_{k-1})\big)n\Big| \\
&< \delta n + n^\beta + C\exp\big(C(p_m+3\theta)n^{\alpha\gamma}\big) =: r(n).
\end{aligned} \tag{79}
$$

Thus, using (78), (79), we have

$$
sn - r(n) < \sum_{i\in I}\phi_k(X_i) \le \Big[\max_{i\in I}\frac{\phi_k(X_i)}{\phi_m(X_i)}\Big]\cdot\sum_{i\in I}\phi_m(X_i) \le \Big[\max_{i\in I}\frac{\phi_k(X_i)}{\phi_m(X_i)}\Big]\cdot(s+\eta)^\gamma n^\gamma,
$$

which implies that for some index i, $\phi_k(X_i)/\phi_m(X_i) \ge (sn - r(n))/(s+\eta)^\gamma n^\gamma)$. Thus, combining this with the condition (26), for sufficiently large n, $M_n^{1-\gamma} \ge \frac{s-2\delta}{(s+\eta)^\gamma}$, since $\lim_{n\to\infty}\frac{r(n)}{n} = \delta$ (recall that $p_m + 3\theta < 0$). This concludes the proof of (77).

Therefore, using (73), (74), (76) and (77), for each $\delta > 0$,

$$
\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{n^\gamma}\log\mathbb{Q}\Big(\Big\{M_n < \Big(\frac{s-2\delta}{(s+\eta)^\gamma}\Big)^{\frac{1}{1-\gamma}}\Big\}\cap C_n^\delta\Big) \\
&\quad\le \limsup_{n\to\infty}\frac{1}{n^\gamma}\log\mathbb{Q}\big(E_n^1\cup E_n^2\cup E_n^3\big) \le (p_m+2\theta)(s+\eta)^\gamma
\end{aligned} \tag{80}
$$

(recall that due to the condition (69), $2\beta - 1 - 2\alpha > \gamma$). Note that due to the condition (75), for sufficiently small $\delta > 0$,

$$
s - \epsilon < \Big(\frac{s-2\delta}{(s+\eta)^\gamma}\Big)^{\frac{1}{1-\gamma}}.
$$

Thus, using (71) and (80), for such $\eta > 0$,

$$
\begin{aligned}
&\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n^\gamma}\log\mathbb{Q}\big(M_n < s-\epsilon \,|\, C_n^\delta\big) \\
&\quad\le \limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n^\gamma}\log\mathbb{Q}\Big(\Big\{M_n < \Big(\frac{s-2\delta}{(s+\eta)^\gamma}\Big)^{\frac{1}{1-\gamma}}\Big\}\cap C_n^\delta\Big) \\
&\qquad - \liminf_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n^\gamma}\log\mathbb{Q}\big(C_n^\delta\big) \\
&\quad\le (p_m+2\theta)(s+\eta)^\gamma - p_m s^\gamma.
\end{aligned} \tag{81}
$$

Since for sufficiently small $\theta > 0$, $(p_m+2\theta)(s+\eta)^\gamma - p_m s^\gamma < 0$ (recall that $p_m + 2\theta < 0$), proof of (27) is concluded.

Now, let us prove (28). Recall that we have the upper tail estimate

$$
\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{Q}\big(M_n \ge a_k - g_2(a_1,\dots,a_{k-1}) + \epsilon \,|\, C_n^\delta\big) < 0 \tag{82}
$$

according to Theorem 2.6. Indeed, changing the reference measure from $\mathbb{P}$ to $\mathbb{Q}$ does not affect the estimate (82) due to the observation Remark 2.7. Combining (82) with (27), (28) immediately follows. □

Theorem 2.8 claims that when $(a_1,\dots,a_{k-1}) \in \mathcal{S}_1$ and $a_k > g_2(a_1,\dots,a_{k-1})$, localization happens in the sense that (28) holds. One can also show that localization only happens at the single site. Let us denote N_n by the second largest component among $\frac{\phi_k(X_i)}{n}$'s, and prove that N_n gets closer to zero in the following sense.

THEOREM 4.2. *Under the same condition as in Theorem 2.8, for* $\epsilon > 0$,

$$\lim_{\delta\to 0}\lim_{n\to\infty}\mathbb{Q}\big(\big\{\big|M_n-\big(a_k-g_2(a_1,\dots,a_{k-1})\big)\big|<\epsilon\big\}\cap\big\{|N_n|<\epsilon\big\}|C_n^\delta\big)=1. \tag{83}$$

PROOF. Throughout the proof, we use the notation $s := a_k - g_2(a_1,\dots,a_{k-1})$,

$$S_{n-2}^i := \frac{\phi_i(X_3)+\cdots+\phi_i(X_n)}{n-2}$$

for each $1 \le i \le k$. In order to prove (83), it suffices to prove that

$$\lim_{\delta\to 0}\lim_{n\to\infty}\mathbb{Q}\big(\big\{|M_n-s|<\epsilon\big\}\cap\big\{|N_n|\le 2\epsilon\big\}|C_n^\delta\big)=1.$$

Thanks to Theorem 2.8, it reduces to show that

$$\lim_{\delta\to 0}\lim_{n\to\infty}\mathbb{Q}\big(\big\{|M_n-s|<\epsilon\big\}\cap\big\{|N_n|> 2\epsilon\big\}|C_n^\delta\big)=0.$$

Thus, proof is concluded once we show the stronger statement

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{Q}\big(\big\{|M_n-s|<\epsilon\big\}\cap\big\{|N_n|> 2\epsilon\big\}|C_n^\delta\big)<0. \tag{84}$$

According to Remark 2.7, it suffices to prove the estimate (84) under the reference measure $\mathbb{P}$ instead of $\mathbb{Q}$. One can take closed sets F_n^δ such that for sufficiently large n,

$$F_n^\delta \subset \prod_{i=1}^{k-1}[a_i-2\delta,a_i+2\delta]\times[0,a_k-s-\epsilon+2\delta]$$

and

$$\left\{\left|\frac{\phi_k(X_1)}{n}-s\right|<\epsilon\right\}\cap\left\{\frac{\phi_k(X_2)}{n}>2\epsilon\right\}\cap C_n^\delta \quad\Rightarrow\quad \big(S_{n-2}^1,\dots,S_{n-2}^k\big)\in F_n^\delta.$$

Since the sequence of empirical means $(S_n^1,\dots,S_n^k)$ satisfy the weak LDP with a rate function I, we have

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(C_n^\delta\big)\ge -\inf_{v_i\in(a_i-\frac{\delta}{2},a_i+\frac{\delta}{2})}I(v_1,\dots,v_k)\ge -I(a_1,\dots,a_k)$$

and

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(\big(S_{n-2}^1,\dots,S_{n-2}^k\big)\in F_n^\delta\big)\\ &\quad\le -\inf_{v_1\in[a_1-2\delta,a_1+2\delta],\dots,v_{k-1}\in[a_{k-1}-2\delta,a_{k-1}+2\delta],v_k\in[0,a_k-s-\epsilon+2\delta]}I(v_1,\dots,v_k).\end{aligned}$$

Sending $\delta\to 0$, using [3], Lemma 4.1.6, and Proposition 3.14,

$$\begin{aligned}&\lim_{\delta\to 0}\inf_{v_1\in[a_1-2\delta,a_1+2\delta],\dots,v_{k-1}\in[a_{k-1}-2\delta,a_{k-1}+2\delta],v_k\in[0,a_k-s-\epsilon+2\delta]}I(v_1,\dots,v_k)\\ &\quad=\inf_{v_k\in[0,a_k-s-\epsilon]}I(a_1,\dots,a_{k-1},v_k)\\ &\quad=I\big(a_1,\dots,a_{k-1},g_2(a_1,\dots,a_{k-1})-\epsilon\big)>I(a_1,\dots,a_{k-1},a_k).\end{aligned}$$

Therefore, combining previous estimates together,

$$\begin{aligned}&\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\left\{\left|\frac{\phi_k(X_1)}{n}-s\right|<\epsilon\right\}\cap\left\{\frac{\phi_k(X_2)}{n}>2\epsilon\right\}|C_n^\delta\right)\\ &\quad\le\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(\big(S_{n-2}^1,\dots,S_{n-2}^k\big)\in F_n^\delta\big)-\liminf_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(C_n^\delta\big)<0.\end{aligned}$$

This concludes the proof of (84). □

5. Examples. In this section, we present some concrete examples of the microcanonical distributions for which the aforementioned theories can be applied. In particular, we establish the principle of equivalence of ensembles, and study the localization and delocalization phenomena.

5.1. *Single constraint.* We first consider the microcanonical ensemble given by a single constraint with an unbounded macroscopic observable. We refer to [3], Section 7.3, for the equivalence of ensembles result for this case. In this section, using the large deviation results obtained in Section 3, we derive the equivalence of ensembles result in a different way. We also prove that localization cannot happen.

Suppose that a function $\phi : (0, \infty) \to (0, \infty)$ satisfies the conditions (i) and (ii) in Assumption 1. Define λ to be a probability measure on $(0, \infty)$ whose distribution is given by $\frac{1}{Z} e^{-\phi}\, dx$. The reference measure on the configuration space $(0, \infty)^{\mathbb{N}}$ is given by $\mathbb{P} = \lambda^{\otimes \mathbb{N}}$, and let us denote $X_i : \Omega \to (0, \infty)$ by the projection onto the ith coordinate. Let us consider the microcanonical ensemble $\mathbb{P}((X_1, \dots, X_n) \in \cdot | C_n^\delta)$, where the constraint is given by

$$C_n^\delta := \left\{ \left| \frac{\phi(X_1) + \cdots + \phi(X_n)}{n} - a \right| \le \delta \right\}. \tag{85}$$

We define $S_n := \frac{\phi(X_1)+\cdots+\phi(X_n)}{n}$ and $H(p) := \log \int e^{p\phi}\, d\lambda$. Note that $H(p) < \infty$ if and only if $p < 1$, and H is differentiable on the interval $(-\infty, 1)$. Thanks to the Cramér's theorem, the sequence S_n under the reference measure $\mathbb{P}$ satisfies the (full) LDP with a good rate function I which is the Legendre transform of H.

Throughout this section, we assume that a belongs to the image of $(-\infty, 1)$ under the map H' in order that the conditional distribution is well defined. In fact, if $a = H'(p)$ for some $p \in (-\infty, 1)$, then $p \in \partial I(a)$, which implies that $I(a) < \infty$. We first derive the equivalence of ensembles result.

PROPOSITION 5.1. *For any fixed positive integer j,*

$$\lim_{\delta \to 0} \lim_{n \to \infty} \mathbb{P}\big((X_1, \dots, X_j) \in \cdot | C_n^\delta\big) = \big(\lambda^*\big)^{\otimes j}. \tag{86}$$

Here λ^* is a probability measure on $(0, \infty)$ whose distribution is given by $\frac{1}{Z} e^{p\phi}\, d\lambda$ for $p \in (-\infty, 1)$ satisfying $H'(p) = a$, or equivalently $\int \phi\, d\lambda^* = a$.

PROOF. Uniqueness of $p \in (-\infty, 1)$ satisfying $H'(p) = a$ is obvious since H is strictly convex on $(-\infty, 1)$. According to the LDP result for the single constraint case (see Remark 3.5) and the Gibbs conditioning principle, (86) holds for λ^* which is a unique minimizer of

$$\mu \mapsto H(\mu|\lambda) + a - \int \phi\, d\mu \tag{87}$$

over the constraint $\int \phi\, d\mu \le a$. For any $\mu \ll dx$ with $\int \phi\, d\mu \le a$,

$$\begin{aligned} H(\mu|\lambda) + a - \int \phi\, d\mu &= H\big(\mu|\lambda^*\big) + p \int \phi\, d\mu + a - \int \phi\, d\mu + C \\ &\ge a - (1-p)a + C \end{aligned} \tag{88}$$

for some universal constant C. Also, equality holds if and only if $\mu = \lambda^*$ since $\int \phi\, d\lambda^* = a$. Thus, the infimum of (87) is uniquely obtained at $\mu = \lambda^*$. This concludes the proof.

Note that in the view of (88), since $H(\mu|\lambda)+a-\int\phi\,d\mu=-h(\mu)+a$, one can also check that

$$\lambda^*=\underset{\int\phi\,d\mu\le a}{\arg\max}\,h(\mu)=\underset{\int\phi\,d\mu=a}{\arg\max}\,h(\mu). \tag{89}$$

□

As in Remark 3.10, Proposition (5.1) holds under the uniform distribution on the constraint C_n^δ as well. Proposition (5.1) claims that in the equivalence of ensembles viewpoint, when we consider the uniform distribution on the single constraint (85) with an unbounded function ϕ, it behaves similarly to the case when ϕ is bounded (see Theorem 3.3 for bounded ϕ). This is a striking difference from the multiple constraints case we have discussed so far.

Now, we show that localization cannot happen when the microcanonical ensemble is given by a single constraint (85).

PROPOSITION 5.2. *For any* $\epsilon>0$,

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(M_n\ge\epsilon|C_n^\delta\big)<0.$$

In particular, localization does not happen in the sense that

$$\lim_{\delta\to 0}\lim_{n\to\infty}\mathbb{P}\big(M_n<\epsilon|C_n^\delta\big)=1.$$

PROOF. Let us choose a reference measure $d\nu=\frac{1}{Z}e^{c\phi}\,d\lambda$ for $c<1$ such that $\int\phi\,d\nu>a$. In fact, such c exists since $\lim_{p\to 1^-}H(p)=\infty$ and H is strictly convex. According to the Cramér's theorem, the sequence S_n under the new reference measure $\mathbb{Q}:=\nu^{\otimes\mathbb{N}}$ satisfies the (full) LDP with a good rate function $\bar{I}(v)$ which is the Legendre transform of $\bar{H}(p)=\log\int e^{px}\,d\nu(x)$. Thus, for each $\delta>0$,

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{Q}\big(C_n^\delta\big)\ge-\inf_{v\in(a-\delta,a+\delta)}\bar{I}(v)\ge-\bar{I}(a). \tag{90}$$

For sufficiently large n, we have

$$\left\{\frac{\phi(X_1)}{n}\in[\epsilon,a]\right\}\cap C_n^\delta\quad\Rightarrow\quad S_{n-1}:=\frac{\phi(X_2)+\cdots+\phi(X_n)}{n-1}\in[0,a-\epsilon+2\delta].$$

Using the fact that $\mathbb{Q}(M_n\in[\epsilon,a])\le n\mathbb{Q}(\frac{\phi(X_1)}{n}\in[\epsilon,a])$, we have

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{Q}\big(\{M_n\in[\epsilon,a]\}\cap C_n^\delta\big)\\
&\quad\le\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{Q}\left(\left\{\frac{\phi(X_1)}{n}\in[\epsilon,a]\right\}\cap C_n^\delta\right)\\
&\quad\le\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{Q}\big(S_{n-1}\in[0,a-\epsilon+2\delta]\big)\le-\inf_{v\in[0,a-\epsilon+2\delta]}\bar{I}(v).
\end{aligned} \tag{91}$$

Sending $\delta\to 0$, applying [3], Lemma 4.1.6, we have

$$\lim_{\delta\to 0}\inf_{v\in[0,a-\epsilon+2\delta]}\bar{I}(v)=\inf_{v\in[0,a-\epsilon]}\bar{I}(v). \tag{92}$$

Now, let us prove that

$$\inf_{v\in[0,a-\epsilon]}\bar{I}(v)>\bar{I}(a). \tag{93}$$

According to [3], Lemma 2.2.5, $\bar{I}$ is nonincreasing on the interval $(0, \int \phi \, d\nu)$. Since $\int \phi \, d\nu > a$, this implies that $\inf_{v \in [0,a-\epsilon]} \bar{I}(v) = \bar{I}(a-\epsilon)$. If $\bar{I}(a-\epsilon) = \bar{I}(a)$, then $\bar{I}(v) = \bar{I}(a)$ for all $v \in (a-\epsilon, a)$, which means that $\bar{I}'(v) = 0$. Thus, $v \in \partial \bar{H}(0)$ for all $v \in (a-\epsilon, a)$, which leads to the contradiction since $\bar{H}$ is differentiable at 0. Thus, $\bar{I}(a-\epsilon) \neq \bar{I}(a)$, and since $\bar{I}$ is nonincreasing on the interval $(0, \int \phi \, d\nu)$, (93) is proved.

Therefore, using (90), (91), (92), and (93),

$$
\begin{aligned}
&\limsup_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{Q}\big(M_n \in [\epsilon, a] | C_n^\delta\big) \\
&\quad \leq \limsup_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{Q}\big(\{M_n \in [\epsilon, a]\} \cap C_n^\delta\big) - \liminf_{\delta \to 0} \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{Q}\big(C_n^\delta\big) \\
&\quad \leq - \inf_{v \in [0,a-\epsilon]} \bar{I}(v) + \bar{I}(a) < 0.
\end{aligned}
$$

Note that according to (24),

$$
\limsup_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\big(A | C_n^\delta\big) = \limsup_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{Q}\big(A | C_n^\delta\big)
$$

for any Borel set A. Therefore, the proof is concluded. □

5.2. *Two constraints*: *l^p spheres*. In this section, we consider the microcanonical distribution given by two l^p-constraints. In particular, we consider the case $\phi_1(x) = x$ and $\phi_2(x) = x^2$. This type of the microcanonical ensemble was previously studied by Chatterjee [2]. He established the convergence of finite marginal distributions and the localization phenomenon. However, the approach used in [2] is ad hoc and only adapted to the special case, so in this section we obtain the result using the unifying theory developed throughout this paper.

It is obvious that $\phi_1(x) = x$ and $\phi_2(x) = x^2$ satisfy Assumption 1. Note that the reference measure $\mathbb{P}$ on the configuration space $(0, \infty)^{\mathbb{N}}$ is given by $\mathbb{P} = \exp(1)^{\otimes \mathbb{N}}$. Since $\int x^2 \, d\mu \geq (\int x \, d\mu)^2$ for any $\mu \in \mathcal{M}_1(\mathbb{R}^+)$ and

$$
\left\{\mu \in \mathcal{M}_1(\mathbb{R}^+) : \mu \ll dx, \int x d\mu = v_1, \int x^2 \, d\mu \leq v_2 \right\}
$$

is a nonempty set whenever $v_2 > v_1^2$, we have $g_1(v_1) = v_1^2$. This means that $\mathcal{A}_1 = \mathbb{R}^+$, and the admissible set is defined by

$$
\mathcal{A} = \{(v_1, v_2) \in (0, \infty)^2 : v_2 > v_1^2\}.
$$

We have $v_1 \in \pi_1(\partial H(p_1, 0))$ for p_1 satisfying $\frac{\int x e^{p_1 x} \, d\lambda}{\int e^{p_1 x} \, d\lambda} = v_1$ (see Definition 2.3 for the meaning of projection π_1). For such p_1 ($p_1 = 1 - \frac{1}{v_1}$), one can check that $\partial H(p_1, 0) = \{(v_1, v_2) \in (0, \infty)^2 | v_2 \geq 2v_1^2\}$ using the fact that $\frac{\int x^2 e^{p_1 x} \, d\lambda}{\int e^{p_1 x} \, d\lambda} = 2v_1^2$. Thus, g_2 can be chosen as $g_2(v_1) = 2v_1^2$. Obviously, $\mathcal{S}_1 = \mathbb{R}^+$ and $\mathcal{S}_2$ is an empty set. Also, according to Proposition 3.14, a weak LDP rate function I for the sequence (S_n^1, S_n^2) satisfies that for any $c > 0$,

$$
I\big(v_1, 2v_1^2\big) = I\big(v_1, 2v_1^2 + c\big). \tag{94}
$$

For $r^2 < s < 2r^2$, define $G_{r,s}$ by a probability measure on $(0, \infty)$ whose distribution is of the form $\frac{1}{Z_{r,s}} e^{\alpha x + \beta x^2} \, dx$ and satisfying

$$
\int x \, dG_{r,s} = r, \qquad \int x^2 \, dG_{r,s} = s. \tag{95}
$$

The existence of such measure can be deduced from Proposition 3.13. We first derive the following equivalence of ensembles result as an application of Theorem 2.4.

PROPOSITION 5.3. *Fix any positive integer* j. *When* $a_1^2 < a_2 < 2a_1^2$,

$$\lim_{\delta\to 0}\lim_{n\to\infty}\mathbb{P}\big((X_1,\dots,X_j)\in\cdot|C_n^\delta\big)=G_{a_1,a_2}^{\otimes j}.$$

On the other hand, *in the case of* $a_2 \ge 2a_1^2$,

$$\lim_{\delta\to 0}\lim_{n\to\infty}\mathbb{P}\big((X_1,\dots,X_j)\in\cdot|C_n^\delta\big)=\exp(a_1)^{\otimes j}.$$

Finally, let us derive the localization and delocalization result. Let us denote M_n by the maximum component $M_n := \max_i \frac{X_i^2}{n}$. Since $\int x\,d\lambda = 1$ and ϕ_1, ϕ_2 satisfy the condition (26), the results of Theorems 2.6 and 2.8 read as follows.

PROPOSITION 5.4. *Suppose that* $a_1 = 1$, *and fix any* $\epsilon > 0$. *In the case of* $1 < a_2 \le 2$, *localization does not happen in the sense that*

$$\lim_{\delta\to 0}\lim_{n\to\infty}\mathbb{P}\big(M_n > \epsilon|C_n^\delta\big)=0.$$

On the other hand, *in the case of* $a_2 > 2$, *localization happens in that*

$$\lim_{\delta\to 0}\lim_{n\to\infty}\mathbb{P}\big(\big|M_n-(a_2-2)\big|>\epsilon|C_n^\delta\big)=0.$$

Note that when $a_2 > 2$, the upper tail estimate for M_n (18) reads as

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(M_n\ge a_2-2+\epsilon|C_n^\delta\big)<0,$$

and the lower tail estimate for M_n (27) reads as

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{\sqrt{n}}\log\mathbb{P}\big(M_n< a_2-2-\epsilon|C_n^\delta\big)<0$$

since $\gamma_1 = \frac{1}{2}$. As explained in Section 2, the maximum component M_n behaves differently in the upper tail and lower tail regime.

5.3. *Three constraints*: l^p *spheres*. The last example we consider is the microcanonical ensemble given by three l^p-constraints. In particular, we assume that $\phi_i(x) = x^i$ for $i = 1, 2, 3$. It is obvious that these functions satisfy Assumption 1. Note that the reference measure $\mathbb{P}$ on the configuration space $(0,\infty)^{\mathbb{N}}$ is given by $\mathbb{P} = \exp(1)^{\otimes\mathbb{N}}$. It is not hard to check that $\mathcal{A}_1 = \{(v_1, v_2) \in (0,\infty)^2 | v_1^2 < v_2\}$, $g_1(v_1, v_2) = \frac{v_2^2}{v_1}$, and the admissible set $\mathcal{A}$ is given by

$$\mathcal{A}=\big\{(v_1,v_2,v_3)\in(0,\infty)^3 : v_1^2<v_2, v_2^2<v_1v_3\big\}.$$

We first characterize the sets S_1 and S_2.

LEMMA 5.5. *The sets* $S_1, S_2 \subset (0,\infty)^2$ *are given by*

$$S_1=\big\{(v_1,v_2): v_1^2<v_2\le 2v_1^2\big\},\qquad S_2=\big\{(v_1,v_2): 2v_1^2<v_2\big\}.$$

PROOF. We first prove the statement for S_1.

Step 1. If $v_1^2 < v_2 \le 2v_1^2$, then there exist p_1, p_2, v_3 such that $(v_1, v_2, v_3) \in \partial H(p_1, p_2, 0)$: first, we claim that there exist p_1, p_2 satisfying that for $i = 1, 2$, $v_i = \frac{1}{Z}\int x^i e^{p_1x+p_2x^2}\,d\lambda$ (Z is a normalizing constant $Z = \int e^{p_1x+p_2x^2}\,d\lambda$). In fact, when $v_1^2 < v_2 < 2v_1^2$, this is proved in Section 5.2, and when $v_2 = 2v_1^2$, one can choose $p_1 = 1 - \frac{1}{v_1}$, $p_2 = 0$. Therefore, for $g_2(v_1, v_2)$ defined by

$$g_2(v_1, v_2) = \frac{1}{Z} \int x^3 e^{p_1 x + p_2 x^2} \, d\lambda,$$

we have $\partial H(p_1, p_2, 0) = \{(v_1, v_2, w) : w \geq g_2(v_1, v_2)\}$ according to Lemma 3.12. This concludes the proof of Step 1.

Step 2. If $2v_1^2 < v_2$, then there does not exist p_1, p_2, v_3 such that $(v_1, v_2, v_3) \in \partial H(p_1, p_2, 0)$: suppose that such p_1, p_2, v_3 exist. Then, by Lemma 3.12, for $i = 1, 2$, $v_i = \frac{1}{Z} \int x^i e^{p_1 x + p_2 x^2} \, d\lambda$. We have $p_2 < 0$ since $v_2 = 2v_1^2$ if $p_2 = 0$. This implies that the logarithmic moment generating function $H_1(p_1, p_2) = \log \int e^{p_1 x + p_2 x^2} \, d\lambda$ is differentiable at (p_1, p_2), and $(v_1, v_2) \in \partial H_1(p_1, p_2)$. By the Legendre duality, $(p_1, p_2) \in \partial I_1(v_1, v_2)$, where I_1 is a Legendre dual of H_1. However, due to (94), $p_2 = 0$ since $v_2 > 2v_1^2$, which leads to the contradiction.

The statement for S_2 is obvious since $S_2 = \mathcal{A}_1 \cap S_1^c$. $\square$

As an application of Theorem 2.4 and Lemma 5.5, we can deduce the following equivalence of ensembles result.

PROPOSITION 5.6. *Fix any positive integer j. Then,*

$$\lim_{\delta \to 0} \lim_{n \to \infty} \mathbb{P}\big((X_1, \ldots, X_j) \in \cdot | C_n^\delta\big) = (\lambda^*)^{\otimes j},$$

where λ^* is characterized as follows: in the case of $a_1^2 < a_2 \leq 2a_1^2$ and $a_3 \geq g_2(a_1, a_2)$,

$$\lambda^* = \frac{1}{Z} e^{p_1 x + p_2 x^2} \, dx$$

for p_1, p_2 satisfying $\int x^i \, d\lambda^* = a_i$ for $i = 1, 2$.

On the other hand, either in the case

(i) $2a_1^2 < a_2$ *or*

(ii) $a_1^2 < a_2 \leq 2a_1^2$ *and* $a_3 < g_2(a_1, a_2)$,

$$\lambda^* = \frac{1}{Z} e^{p_1 x + p_2 x^2 + p_3 x^3} \, dx$$

for p_1, p_2, p_3 satisfying $p_3 < 0$ and $\int x^i \, d\lambda^* = a_i$ for $i = 1, 2, 3$.

Finally, since ϕ_i's satisfy the condition (26), one can derive the localization and delocalization result as applications of Theorems 2.6 and 2.8.

PROPOSITION 5.7. *Suppose that $(a_1, a_2) \in \mathcal{S}_2$. Then, localization does not happen in the sense that*

$$\lim_{\delta \to 0} \lim_{n \to \infty} \mathbb{P}\big(M_n > \epsilon | C_n^\delta\big) = 0. \tag{96}$$

On the other hand, assume that $(a_1, a_2) \in \mathcal{S}_1$. In the case of $a_3 \leq g_2(a_1, a_2)$, localization does not happen in the sense that (96) holds. However, in the case of $a_3 > g_2(a_1, a_2)$, under the reference measure $\mathbb{Q} = \nu^{\otimes 3}$ with ν of the form

$$\nu = \frac{1}{Z} e^{p_1 x + p_2 x^2} \, dx, \tag{97}$$

satisfying $\int x^i \, d\nu = a_i$ for $i = 1, 2$, localization happens in the sense that

$$\lim_{\delta \to 0} \lim_{n \to \infty} \mathbb{Q}\big(\big|M_n - \big(a_3 - g_2(a_1, a_2)\big)\big| > \epsilon | C_n^\delta\big) = 0.$$

Note that when $a_3 > g_2(a_1, a_2)$, the upper tail estimate for M_n (18) reads as

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{Q}\big(M_n \ge a_3 - g_2(a_1,a_2)+\epsilon|C_n^\delta\big) < 0,$$

and the lower tail estimate for M_n (27) reads as

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n^\gamma}\log\mathbb{Q}\big(M_n < a_3 - g_2(a_1,a_2)-\epsilon|C_n^\delta\big) < 0.$$

Here, $\gamma = \frac{1}{3}$ when $p_2 = 0$ in the expression (97), and $\gamma = \frac{2}{3}$ when $p_2 < 0$ in the expression (97), since $\gamma_i = \frac{i}{3}$ for $i = 1, 2$.

APPENDIX: AUXILIARY LEMMA

We prove the following auxiliary lemma frequently used in the paper.

LEMMA A.1. *Suppose that Assumption 1 holds. Also, for some* $1 \le m \le k-1$, *consider the probability distribution* $\nu = \frac{1}{Z}e^{p_1\phi_1+\cdots+p_m\phi_m}\,dx$ *on* $(0,\infty)$ *with* $p_m < 0$. *Then, for any number* $M \ge 0$ *and* $\epsilon > 0$,

$$\lim_{n\to\infty}\frac{1}{n}\log\nu\left(\left|\frac{\phi_k(X_1)}{n} - M\right| < \epsilon\right) = 0. \tag{98}$$

Let us denote $\mathbb{Q}$ *by the product measure* $\mathbb{Q} = \nu^{\otimes\mathbb{N}}$. *Then, for any* $0 < \theta < -p_m$, *there exists* $C = C(\theta) > 0$ *such that*

$$\mathbb{Q}\left(\sum_{i=1}^{j}\phi_m(X_i) > M\right) < C(M - Cj)^{j-1}\exp\big[(p_m+\theta)(M - Cj)\big] \tag{99}$$

for any $j \in \mathbb{N}$, $M > Cj + 2$.

Furthermore, under the additional condition (26),

$$\liminf_{n\to\infty}\frac{1}{n^{\gamma_m}}\log\nu\left(\left|\frac{\phi_k(X_1)}{n} - M\right| < \epsilon\right) \ge p_m M^{\gamma_m}. \tag{100}$$

PROOF. Note that due to Assumption 1, for any $\theta > 0$, there exists $C = C(\theta)$ such that

$$x > C \quad\Rightarrow\quad (p_m - \theta)\phi_m < p_1\phi_1 + \cdots + p_m\phi_m < (p_m+\theta)\phi_m.$$

Let us first prove (98). Since $m < k$, thanks to the condition (iii) in Assumption 1, there exists $0 < \delta < 1$ such that for sufficiently large y, $\sum_{i=1}^m p_i\phi_i(\phi_k^{-1}(y)) < (p_m+\delta)y^{1-\delta}$. Thus, using the condition (C4) in Assumption 1 and the change of variables, for sufficiently large n,

$$\begin{aligned}\nu\left(\left|\frac{\phi_k(X_1)}{n} - M\right| < \epsilon\right) &= \int_{(M-\epsilon)n}^{(M+\epsilon)n} e^{\sum_{i=1}^m p_i\phi_i(\phi_k^{-1}(y))}\frac{1}{\phi_k'(\phi_k^{-1}(y))}\,dy \\ &< \int_{(M-\epsilon)n}^{(M+\epsilon)n} Ce^{(p_i+\delta)y^{1-\delta}}y^C\,dy \\ &< C\epsilon n e^{(p_i+\delta)((M+\epsilon)n)^{1-\delta}}\big((M+\epsilon)n\big)^C.\end{aligned}$$

After taking log and dividing by n, and then sending $n\to\infty$, we obtain (98).

Let us now prove (99). If we define $Y_i := \phi_m(X_i)$, then Y_i's are i.i.d. whose individual distribution is given by $\frac{1}{Z}e^{\sum_{i=1}^m p_i\phi_i(\phi_m^{-1}(y))}\frac{1}{\phi_m'(\phi_m^{-1}(y))}\,dy$ on $(0,\infty)$. Using Assumption 1, for any $0 < \theta < -p_m$, there exists C such that

$$y > C \quad\Rightarrow\quad \frac{1}{Z}e^{\sum_{i=1}^m p_i\phi_i(\phi_m^{-1}(y))}\frac{1}{\phi_m'(\phi_m^{-1}(y))} < \frac{1}{Z'}e^{(p_m+\theta)y} \tag{101}$$

($Z' = \int e^{(p_m+\theta)y}\,dy$ is a normalizing constant). Let us denote $Z_1, Z_2, \dots$ by i.i.d. random variables whose individual distribution is given by $\exp(p_m + \theta)$. Then, (101) implies that for any $K > 0$,

$$\mathbb{Q}\left(\sum_{i=1}^{j} \phi_m(X_i)\mathbb{1}_{\phi_m(X_i)\geq C} > K\right) \leq \mathbb{Q}\left(\sum_{i=1}^{j} Z_i > K\right) \tag{102}$$

by the simple coupling argument. Using the fact that law of $\sum_{i=1}^{j} Z_i$ is Gamma$(j, p_m + \theta)$, it is easy to check that for $K > 2$,

$$\mathbb{Q}\left(\sum_{i=1}^{j} Z_i > K\right) < CK^{j-1}e^{(p_m+\theta)K} \tag{103}$$

(we refer to [2] for the estimate (103) in the case Gamma$(j, 1)$ distribution). On the other hand, it is obvious that

$$\sum_{i=1}^{j} \phi_m(X_i) > M \quad \Rightarrow \quad \sum_{i=1}^{j} \phi_m(X_i)\mathbb{1}_{\phi_m(X_i)\geq C} > M - Cj. \tag{104}$$

Thus, (102), (103), and (104) conclude the proof of (99).

Finally, let us prove (100) under the additional condition (26). Using Assumption 1, condition (26), and the change of variables, for any $\theta, \eta > 0$,

$$\nu\left(\left|\frac{\phi_k(X_1)}{n} - M\right| < \eta\right) > C\eta n e^{(p_m-\theta)((M+\eta)n)^{\gamma_m}}\left((M-\eta)n\right)^{C}$$

for sufficiently large n. After taking log, dividing by n, sending $n \to \infty$, and then sending $\theta \to 0$, we have

$$\liminf_{n\to\infty} \frac{1}{n^{\gamma_m}} \log \nu\left(\left|\frac{\phi_k(X_1)}{n} - M\right| < \eta\right) \geq p_m(M+\eta)^{\gamma_m}.$$

Since for $0 < \eta < \epsilon$,

$$\liminf_{n\to\infty} \frac{1}{n^{\gamma_m}} \log \nu\left(\left|\frac{\phi_k(X_1)}{n} - M\right| < \epsilon\right) > \liminf_{n\to\infty} \frac{1}{n^{\gamma_m}} \log \nu\left(\left|\frac{\phi_k(X_1)}{n} - M\right| < \eta\right),$$

and $\eta > 0$ can be arbitrary small, we obtain (100). $\square$

Acknowledgments. The authors would like to thank anonymous referees for their constructive comments. The author also thanks to Professor Fraydoun Rezakhanlou for suggesting this project and sharing numerous valuable discussions.

REFERENCES

[1] CHATTERJEE, S. (2014). Invariant measures and the soliton resolution conjecture. *Comm. Pure Appl. Math.* **67** 1737–1842. MR3263670 https://doi.org/10.1002/cpa.21501

[2] CHATTERJEE, S. (2017). A note about the uniform distribution on the intersection of a simplex and a sphere. *J. Topol. Anal.* **9** 717–738. MR3684622 https://doi.org/10.1142/S1793525317500224

[3] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **38**. Springer, New York. MR1619036 https://doi.org/10.1007/978-1-4612-5320-4

[4] DEUSCHEL, J.-D., STROOCK, D. W. and ZESSIN, H. (1991). Microcanonical distributions for lattice gases. *Comm. Math. Phys.* **139** 83–101. MR1116411

[5] ELLIS, R. S. (1985). *Entropy, Large Deviations, and Statistical Mechanics. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **271**. Springer, New York. MR0793553 https://doi.org/10.1007/978-1-4613-8533-2

[6] GEORGII, H.-O. (1993). Large deviations and maximum entropy principle for interacting random fields on $\mathbf{Z}^d$. *Ann. Probab.* **21** 1845–1875. MR1245292

[7] GEORGII, H.-O. (1994). Large deviations and the equivalence of ensembles for Gibbsian particle systems with superstable interaction. *Probab. Theory Related Fields* **99** 171–195. MR1278881 https://doi.org/10.1007/BF01199021

[8] GEORGII, H.-O. (1995). The equivalence of ensembles for classical systems of particles. *J. Stat. Phys.* **80** 1341–1378. MR1349785 https://doi.org/10.1007/BF02179874

[9] GEORGII, H.-O. (2011). *Gibbs Measures and Phase Transitions*, 2nd ed. *De Gruyter Studies in Mathematics* **9**. de Gruyter, Berlin. MR2807681 https://doi.org/10.1515/9783110250329

[10] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363

[11] KIM, S. S. and RAMANAN, K. (2018). A conditional limit theorem for high-dimensional ℓ^p-spheres. *J. Appl. Probab.* **55** 1060–1077. MR3899928 https://doi.org/10.1017/jpr.2018.71

[12] LÉONARD, C. (2010). Entropic projections and dominating points. *ESAIM Probab. Stat.* **14** 343–381. MR2795471 https://doi.org/10.1051/ps/2009003

[13] LEWIS, J. T., PFISTER, C.-E. and SULLIVAN, W. G. (1995). Entropy, concentration of probability and conditional limit theorems. *Markov Process. Related Fields* **1** 319–386. MR1403087

[14] RASSOUL-AGHA, F. and SEPPÄLÄINEN, T. (2015). *A Course on Large Deviations with an Introduction to Gibbs Measures*. *Graduate Studies in Mathematics* **162**. Amer. Math. Soc., Providence, RI. MR3309619

[15] ROCKAFELLAR, R. T. (1970). *Convex Analysis*. *Princeton Mathematical Series*, *No*. 28. Princeton Univ. Press, Princeton, NJ. MR0274683

[16] SZNITMAN, A.-S. (1991). Topics in propagation of chaos. In *École D'Été de Probabilités de Saint-Flour XIX—1989*. *Lecture Notes in Math.* **1464** 165–251. Springer, Berlin. MR1108185 https://doi.org/10.1007/BFb0085169

[17] TAO, T. (2006). *Nonlinear Dispersive Equations*: *Local and Global Analysis*. *CBMS Regional Conference Series in Mathematics* **106**. Amer. Math. Soc., Providence, RI. MR2233925 https://doi.org/10.1090/cbms/106